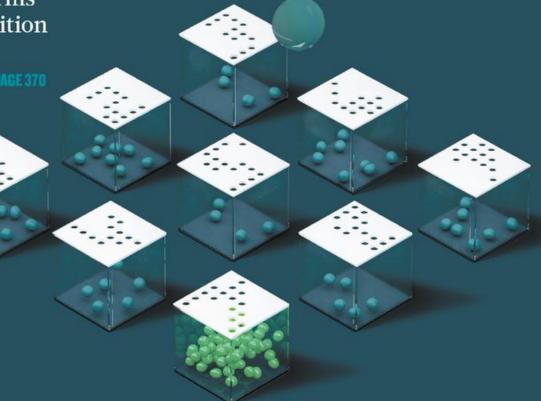
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

PARSED JUDGEMENT

DNA-based neural network performs pattern recognition on 'molecular handwriting'





BOOKS

HOLIDAY READING

A refreshing selection of science for the summer PAGE 328

IMAGING

MOLECULAR RESOLUTION

Microscope sees deeper into the sub-ångström realm PAGES 334 & 343 CAREERS

WINNING STREAK

The secrets of a successful spike in performance

⇒ NATURE.COM/NATURE

19 July 2018 £10



THIS WEEK

EDITORIALS



WORLD VIEW The power of science to change minds in the abortion debate **p.303**

GENOMICS Tracking the lost dogs of the Americas **p.305**

Realism trumps hope at the EPA

Scott Pruitt's resignation from the US Environmental Protection Agency was long overdue. But the threat to science posed by Trump and his allies remains.

he most remarkable thing about Scott Pruitt's resignation from the US Environmental Protection Agency (EPA) is that it took so long. By all accounts, he was unfit to lead one of the world's top science-based regulatory agencies. It wasn't just that the former Oklahoma attorney-general had a well-documented history of consorting with industry to fight the agency. It was his contradictory behaviour — exemplified by the installation of an expensive sound-proof phone booth in his office — which put a premium on secrecy even as he made grand proclamations about transparency. But by far the worst was Pruitt's utter disregard for both the science and the scientists under his charge — as we highlight in a News Feature this week (see page 316).

Ultimately, Pruitt seems to have been felled by the impunity he exhibited over the course of nearly a year and a half at the agency. Lawmakers on both sides of the political aisle raised alarms over his lavish spending and a series of alleged ethical transgressions that are more typical of crony governments elsewhere in the world.

His departure is welcome, but it would be naive to think that the prospects for the agency and its scientists are any brighter. His agenda — the same one as US President Donald Trump — remains intact. Trump made this all too clear in a pair of tweets announcing Pruitt's resignation on 5 July. The president declared that Pruitt had done an "outstanding job", and said that the new acting administrator, Andrew Wheeler, a former coal lobbyist, "will continue on with our great and lasting EPA agenda".

Trump has yet to formally nominate Wheeler as the next EPA administrator, but the move would be in keeping with the president's approach. A lawyer by training, Wheeler spent 4 years at the agency in the early 1990s, under former presidents George H. W. Bush and Bill Clinton. He later served as a top aide on the Senate Environment and Public Works Committee under Oklahoma Republican James Inhofe, a leading climate sceptic in Congress. Wheeler knows how the agency works, and is comfortable on Capitol Hill. In the words of one EPA scientist, who asked for anonymity, Wheeler is "a supremely effective and precise Washington operative". This, of course, is both praise and a warning.

Wheeler will probably restore some kind of normal order at the agency, which means following conventional procedures, reestablishing fractured relations with staff scientists and avoiding the kind of embarrassing headlines that plagued Pruitt's tenure. Already, in his first week as acting administrator, Wheeler has delivered an all-hands address at the agency's research campus in Durham, North Carolina. That stands in stark contrast to Pruitt, who quietly dipped in and out of the campus a few weeks ago, before he stepped down, with no word to the full staff. Not once during his tenure did Pruitt make time to address the EPA's Office of Research and Development, which houses the bulk of the agency's scientists — hardly the way to either inspire loyalty or demonstrate he was on top of his brief.

Under a new boss, EPA researchers might even be able to present

their findings once again to the leadership, as the administration deliberates over environmental and public-health regulations. Such scientific consultations — fundamental to the establishment of science-based policies that can withstand the inevitable legal challenges that follow — were often eschewed under Pruitt, who showed little regard for the importance of evidence.

Scientists should be wary about celebrating Pruitt's exit. They should be careful what they wish for. The problem is that if Wheeler — or whoever takes on the job full-time — is more effective than Pruitt (and they could hardly be otherwise), then Trump's problematic policies are likely to have more impact, too. And that could spell more trouble for public health and the environment, not just in the United States but around the globe — at a time when a sound and evidence-based approach to both has never been so critical.

A fundamental goal of many of Trump's efforts and policies is to relieve US industry of what he regards as regulatory burdens. Republican rhetoric has been trending in that direction for years, par-

"What's needed are policies that allow agency researchers to follow the science wherever it might lead." ticularly when it comes to regulations that combat climate change but that industries find expensive or cumbersome. Indeed, the vast majority of conservative lawmakers have either actively disavowed mainstream science or turned a blind eye to the pressing need to address one of the biggest challenges of the twenty-first century. It's a disgrace that will go down in the history books, but

Trump and his team have pushed things to a new extreme. Rather than simply rolling back regulations, Pruitt sought to straitjacket the EPA and undermine the role of both science and scientists in regulatory policy. For example, he banned scientists with EPA grants from serving on the agency's advisory boards, and proposed a rule that would prevent the agency from citing public-health research for which the underlying data are not publicly available — including high-quality epidemiological studies that help to provide the technical basis for current air-quality regulations, but whose data must be partially hidden to protect patients' identities.

Republicans on Capitol Hill have provided a glimmer of hope by repeatedly rejecting Trump's proposals to slash the EPA budget as well as funding for climate and energy research at other agencies, but money alone won't solve the problems that EPA scientists face today. What's needed are policies that give deference to evidence and that allow agency researchers to follow the science wherever it might lead — even if politicians don't like the implications.

Environmentalism needn't be a partisan issue. It was one of the Republicans' own, Richard Nixon, who oversaw the creation of the EPA, and the last major upgrade to the Clean Air Act came under the first Bush administration. Wheeler might be more successful in implementing Trump's policies, but that's dangerous, as Trump is completely out of touch with scientific reality.

Lost and found

European funders are right to consider the career prospects of young scientists.

't is a century since the first génération perdue came of age. The phrase is attributed to US writer Gertrude Stein, who heard it as a casual insult aimed by a garage boss at a young French mechanic who was working — too slowly — on Stein's car. The term is now generally used to describe a group of people who are lost to society. So when European officials spoke at a conference session last week called 'The lost generation of European scientists', for many participants the name would have conjured up thoughts of an exodus of talented early-career researchers, who are fed up with the insecurity of short-term jobs and with dwindling opportunities in academia. And so it should have: in many disciplines, that issue is real, growing and serious. Young and early-career researchers need the problem to be taken seriously — and so does the rest of the scientific community. Figures are difficult to come by, but less than one-fifth of US postdocs secure a tenured research position, and the situation is even more competitive in Europe.

The 'lost generation' tag has another, more subtle meaning. Popularized by US writer Ernest Hemingway, it was used to describe the age group — Hemingway included — that had been left disoriented and confused by growing up amid the horrors and chaos of the First World War. Lost, not missing. The distinction is important. Careers outside academia are just as valuable and senior scientists must acknowledge this. Nevertheless, young researchers are too often led to believe that a non-academic career is inferior, so individual scientists who find they need to look elsewhere often feel let down, deceived and cynical.

Last week's event, held at the EuroScience Open Forum in Toulouse,

France, covered all of that ground. The session was well attended and was frank about the scale of the problem and the difficulty of finding solutions. This might indicate that European funders and policymakers are catching up with the United States, where the crisis of confidence and opportunity among young scientists — especially in biomedicine — has been widely debated for at least a decade. That would be good news. The bad news is that Europe's fragmented, variable national research bodies and strong university autonomy make it much easier to acknowledge the problem than to change the systems that cause it.

At the meeting, European Research Council president Jean-Pierre Bourguignon hinted at an obvious fix: increase funding for scientific research and create more permanent academic jobs. But that's a big

"Universities should track and provide data on how many academic jobs are available." ask, and one that would take time. A moreimmediate solution calls for more-specific and targeted changes. One is the creation of more full-time staff scientist positions, although such posts (with benefits such as pensions) raise institution costs.

As we have argued previously (*Nature* **550**, 429; 2017), there is a pressing need for

greater transparency about the likelihood of PhD students and post-docs following an academic career to the higher levels. A suggestion made at last week's session — and one that *Nature* endorses — is that universities and other institutions should track and provide data on how many academic jobs are available at each level, and list the destination of every scientist who moves on. The US National Academies has made an attempt at doing this for postdocs, and the European Science Foundation has tried to track the fate of Europe's PhD holders. Both are good models to follow.

Better information won't solve all the problems of all the 'lost' researchers, but it will at least provide them with a map as they decide on their next move. Those who supervise PhD students and postdocs must show them such a map, and take responsibility for preparing them for non-academic careers. What might look like a loss for academia can still be a great gain for society.

Gone rogue

Officials and scientists need help to track down source of a worrying rise in CFC emissions.

fter three decades of being lauded as pioneers in successful environmental stewardship, officials who safeguard Earth's ozone layer are facing an unexpected crisis: how to identify and cut off a rogue new source of ozone-destroying chemicals (S. A. Montzka et al. Nature 557, 413–417; 2018). If not stopped, the emissions of CFC-11 might delay by several decades the healing of ozone holes that appear at high latitudes early each spring. As expected, the issue featured heavily at last week's meeting in Vienna of the Open-Ended Working Group (OEWG) of the Montreal Protocol, which protects the ozone layer. Since the protocol's launch in 1987, countries have curbed the use of ozone-depleting chemicals in refrigeration and other industrial processes.

Ahead of the meeting, media reports and an analysis by the London-based Environmental Investigation Agency — a non-governmental organization with observer status in the Montreal Protocol — used interviews with company executives and information contained in advertisements to suggest that foam-manufacturing companies in rural China are to blame. Chinese delegates in Vienna made it clear that they take the matter seriously and, by all accounts, the issue has gone up to the level of Chinese President Xi Jinping. But they remained extremely reluctant to concede any serious wrongdoings on the part of

Chinese companies, or government negligence in their oversight. This is understandable given that there is not yet definite evidence concerning the sources, quantity, duration or nature of the rogue emissions.

The suspicion that Chinese factories are the main — perhaps the sole — source of the damaging CFC-11 chemicals cannot be dismissed. But for now, increased vigilance must apply to the whole of South and East Asia. To pinpoint the source of the rogue emissions precisely, members of the Montreal Protocol's scientific assessment panel are working to analyse the most recent data from the region's atmospheric monitoring stations, including those of South Korea and Japan. Governments must make available, without delay, any data required for further analysis, and should also provide any other intelligence, such as that from commercial register entries, advertisements or customs, that could help to pin down any source of the emissions. The issue is a test of the strength and muscle of the Montreal Protocol regime, which must mobilize all the pieces — science, monitoring, verification and, possibly, sanctions. Already, four years have elapsed since scientists observed and reported the worrying CFC spike. What's needed now, besides enduring vigilance, is a rapid political and institutional effort.

There is no doubt that China has, over the past few years, stepped up its environmental efforts, including those tackling air pollution and greenhouse-gas emissions. If Chinese sources of CFC-11 production are confirmed, the government should engage its full enforcement capacity to stop it immediately. Ironically, the current crisis is an opportunity for China to demonstrate its emerging leadership in the enforcement of global environmental policies. If the Montreal Protocol survives this test, the most beneficial environmental pact the world has ever undertaken will surely emerge stronger than ever.

SEVEN DAYS The news in brief

RESEARCH

India tests crew pod

India's space agency has successfully tested a crew escape system, which puts the country a step closer to achieving human space flight. A simulated crew module weighing 12.6 tonnes, containing the unoccupied escape system, blasted off from the Satish Dhawan Space Centre in Sriharikota on 5 July. The module was propelled by seven motors that are designed to move crew members away from a threat without exceeding safe g forces. It reached an altitude of nearly 2.7 kilometres before separating from the motors, deploying its parachutes and drifting back to Earth. The module was recovered in the nearby Bay of Bengal.

Sterile mosquitoes

A mosquito-control trial co-sponsored by the Google life-sciences spin-off Verily of South San Francisco, California, has passed one of its first big tests. On 10 July, Australia's funding agency, the Commonwealth Scientific and Industrial Research Organisation, which is another co-sponsor, reported a greater than 80% reduction in dengue-transmitting Aedes aegypti mosquitoes during a five-month trial in three Australian towns in North Oueensland. The trial involved releasing male mosquitoes that had been rendered sterile by infection with Wolbachia bacteria. Sterile males, released into homes, compete for mates with other males, resulting in an eventual population reduction.

Cosmic maps

A European space-telescope project that tracked the faint afterglow of the Big Bang released its final and most



South Africa unveils huge radio telescope

This radio-wave image of the central regions of the Milky Way was unveiled at the opening ceremony of South Africa's MeerKAT radio telescope on 13 July. The 4.4-billion-rand (US\$330-million) observatory has 64 dishes, each 13.5 metres in diameter. It will observe transient astrophysical events — including fast radio bursts — and conduct surveys such as mapping hydrogen abundance across cosmic history. MeerKAT will eventually be part of the

Square Kilometre Array (SKA), a facility under construction in South Africa and Australia, and it is already boosting the country's scientific community and drawing researchers from around the world. "Back in the day, our astronomers went abroad to do astronomy. Now we're the attraction," says Justin Jonas, chief technologist at the Cape Town-based South African Radio Astronomy Observatory. See go.nature.com/2jtg5sd for more.

precise maps of the early Universe on 17 July (see go.nature.com/2jt0sbi). The European Space Agency's Planck telescope launched in 2009 and finished surveying the cosmic microwave background in 2013, but the science team has revamped its data-analysis techniques to improve the precision of its measurements of crucial features of the Universe. The Planck data continue to predict that the current Universe should expand about 9% slower than observations of relatively nearby galaxies suggest. The telescope did not detect the signature of gravitational waves from the early stages of the Big Bang, which would point to an early, exponential expansion known

as inflation, although future,

more-sensitive experiments could yet find such a signature.

EVENTS

Herbicide lawsuits

A US federal judge ruled last week that lawsuits against Monsanto — an agriculture corporation based in St Louis, Missouri — can proceed. One such suit alleges that the company's herbicide, a glyphosate product marketed as Roundup, was a substantial factor in development of the cancer non-Hodgkin's lymphoma in a 42-year-old groundskeeper. Brent Wisner, the plaintiff's lawyer, pointed to evidence linking glyphosate exposure among agricultural workers to cancer. Arguments for this lawsuit began on 9 July, and it is the first of many similar suits filed in the United States to go to trial. "The scientific evidence clearly shows that glyphosate was not the cause" of cancer, wrote Monsanto vice-president Scott Partridge in a statement to Nature.

Indian universities

The Indian government named its first six 'institutes of eminence' on 9 July, with the aim of elevating Indian universities in global rankings. Three public and three private universities will each receive US\$145.7 million over five years and be given greater autonomy to hire foreign faculty members and students. The government was originally due to name 20 institutions of eminence in

SOUTH AFRICAN RADIO ASTRONOMY OBSERVATORY

April. Some academics told *Nature* the delay was because of disagreements on how to select institutions. Others have also criticized the inclusion of the Jio Institute, a private institution that has yet to be set up. "Any institution needs to be in existence and have shown its worth before being considered for institutions of eminence status," says Subhash Lakhotia, a cytogenetics researcher at Banaras Hindu University in Varanasi.

PEOPLE

Terrapin trouble

The US attorney's office indicted David Sommers, a Pennsylvania resident, on 10 July for allegedly poaching thousands of protected terrapins and their eggs from New Jersey salt marshes, and then selling them from his home. An investigation by the US Fish and Wildlife Service found that Sommers sold more than 3,500 diamondback terrapins (Malaclemys terrapin), a turtle species prized in the pet trade for the diamond-shaped markings on its shell, in the United States and Canada. Sommers now awaits trial for allegedly violating state and federal laws.

NASA nominee

US President Donald Trump nominated James Morhard, a Senate staff member, to the number-two job at NASA on



12 July. Morhard (pictured) has degrees in business and law but no space expertise; he is currently serving as the Senate's deputy sergeant at arms. If confirmed, Morhard would join NASA administrator Jim Bridenstine, a former politician who had reportedly been seeking a technical expert as his deputy. In a statement, Bridenstine said that "this administration is committed to American leadership in space".

Mail-fraud deal

A mechanical engineering professor at the University of Colorado Boulder pleaded guilty to mail fraud earlier this month. According to the US attorney's office in Denver, Colorado, Oleg Vasilyev applied for a US\$234,000 federal contract with the US Department of Energy's (DOE's) Los Alamos National Laboratory in New Mexico without informing the university. The funds were transferred to a university account, to which Vasilyev

submitted a series of claims for unallowable expenses, including more than \$140,000 for travel. In a plea agreement dated 6 July, Vasilyev agreed to repay the university \$185,879; the university has reimbursed the DOE for the misappropriated funds. The guilty plea terminated Vasilyev's employment, a university spokesperson said.

POLICY

Innovation council

After Brexit, the United Kingdom will probably not be able to take part in all aspects of the European Union's next major research funding programme, Horizon Europe, EU officials have confirmed. The European Commission's provisional plans for the 7-year programme, which begins in 2021 and has a proposed budget of €100 billion (US\$117 billion), suggested that it would be the most global yet, with many parts of it open to non-EU nations paying a subscription fee. But at the EuroScience Open Forum in Toulouse, France, on 11 July, Robert-Jan Smits, the European Commission's special envoy for open data and former director-general of research, said that the European Innovation Council, a programme launching in 2021, will not be accessible to the United Kingdom and other non-EU 'third'

countries because the council is aimed at supporting start-up companies in the EU. Precise details are still to be negotiated as the EU hammers out the legal framework for third countries in Horizon Europe.

Fossil-fuel vote

Ireland's lower house of parliament voted on 12 July to divest its €8.9-billion (US\$10.4-billion) stateowned investment fund of fossil-fuel companies "as soon as practicable". The upper house must now vote on the measure, but backers say that the proposal has the support of the Irish prime minister and is expected to become law. If that happens, Ireland would become the first country to set a goal of withdrawing all investments in fossil fuels. Norway is also considering whether to divest its \$1-trillion state-owned fund of its holdings in fossil-fuel companies, but the government has yet to come to a decision.

Integrity inquiry

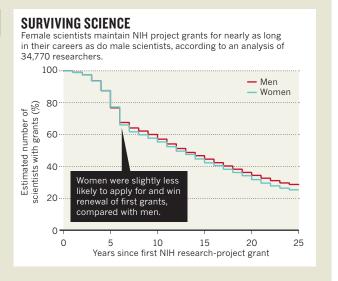
The United Kingdom should establish a committee to monitor universities' misconduct investigations, a parliamentary inquiry recommended on 11 July. The inquiry found that one in four UK universities do not comply with transparency guidelines for such investigations. The report, released by the House of Commons Science and Technology Committee, says there is a need to address the potential conflicts of universities policing themselves. Although the advocacy group Universities UK issued a concordat in 2012 that requires universities to deal with misconduct cases transparently and robustly, there are currently no sanctions for non-compliance. Some universities told the inquiry that they feared disclosing cases of research fraud would undermine their reputation. See go.nature.com/2meydnz for more.

◇ NATURE.COM

For daily news updates see:

TREND WATCH

An analysis of almost 35,000 biomedical researchers has found that after women secure their first major research grant from the US National Institutes of Health (NIH), they are almost as successful as men at netting more NIH funding. The finding casts doubt on the belief that women leave science at a faster rate than men as their careers progress. Women remain underrepresented among NIH grantees, however, winning less than onethird of the grants despite earning nearly half the PhDs in the field.



NEWSINFOCUS

EPIDEMIOLOGY Police in China expand sewage analysis to monitor illegal drug use **p.310**

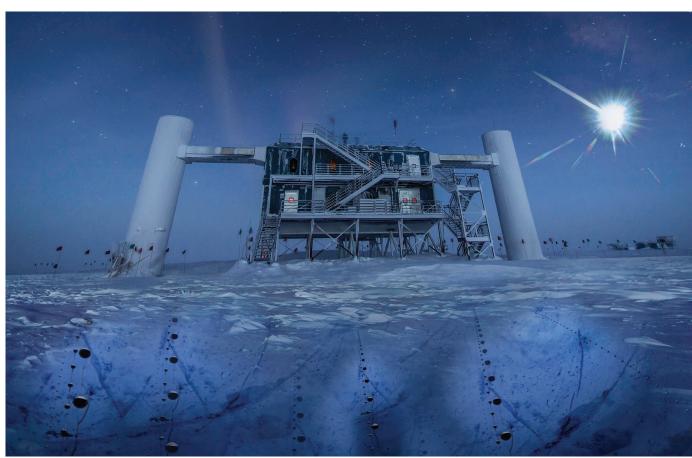
CECUBE/NSF

PUBLISHING Gates Foundation trial ends as journals grapple with open-access issue **p.311**

SPACE Ten new moons of Jupiter help trace Solar System history **p.312**



POLITICS What scientists have faced inside Trump's EPA p. 316



The IceCube lab is in Antarctica, on the South Pole.

ASTROPHYSICS

Particle traced from space

When a neutrino streaked through Antarctica, astrophysicists raced to find the source.

BY DAVIDE CASTELVECCHI

A single subatomic particle detected at the South Pole last September is helping to solve a major cosmic mystery: what creates electrically charged cosmic rays, the most energetic particles in nature.

Follow-up studies on the particle's trajectory by more than a dozen observatories suggest that researchers have, for the first time, identified a distant galaxy as a source of high-energy neutrinos. This discovery could, in turn, help scientists to pin down the still-mysterious

sources of cosmic rays, the protons and atomic nuclei that arrive at Earth from outer space. The same mechanisms that produce cosmic rays should also make high-energy neutrinos.

Multiple teams of researchers from around the world described the neutrino's source in at least seven papers released on 12 July. "Everything points to this as the ultra-bright, energetic source — a gorgeous source," says Elisa Resconi, an astroparticle physicist at the Technical University of Munich in Germany.

Astrophysicists have proposed a number of scenarios for astrophysical phenomena that

could produce both high-energy neutrinos and their electrically charged counterparts, cosmic rays. But until now, they had not managed to unambiguously trace any of these particles back to their source.

MUON ALERT

The story began on 22 September 2017, when an electrically charged particle called a muon zipped through the Antarctic ice cap at close to the speed of light. the IceCube Observatory — an array of more than 5,000 sensors buried in a cubic kilometre's worth of ice — detected

Is flashes of light that the muon produced in its wake. The particle seemed to emerge from below the detector — an orientation that indicated it was the decay product of a neutrino that had come from below the horizon. Muons can travel only so far inside matter, whereas neutrinos often pass through the entire planet unimpeded; most of the muons that IceCube detects originate from neutrinos that have crashed with a particle inside Earth.

Within seconds, a computer cluster at the US National Science Foundation's Amundsen–Scott South Pole Station had reconstructed the precise path of the particle and recognized that the muon had come from a highly energetic neutrino; 43 seconds after the event, the station sent an automated alert to a network of astronomers through a satellite link. It tagged the neutrino as IceCube-170922A.

After receiving the alert, Derek Fox, an astrophysicist at Pennsylvania State University in University Park, quickly secured observing time on the X-ray observatory Swift, which orbits Earth. He and his team found nine sources of high-energy X-rays close to where the neutrino had come from. Among them was an object called TXS 0506+056. This is a blazar, a galaxy with a supermassive black hole at the centre and a known source of γ -rays. In a blazar, the black hole stirs up gas to temperatures of millions of degrees and shoots it out of its poles in two highly collimated jets. In

this case, one of the jets points in the direction of the Solar System. Fox's team announced its findings to the astronomical community the next day.

In the following days, another team inspected data from Fermi-LAT, the Large Area Telescope aboard NASA's Fermi Gammaray Space Telescope. Fermi-LAT constantly sweeps the sky, and among other things monitors about 2,000 blazars. These objects

go through periods of increased activity that can last weeks or months, during which they become unusually bright. "When we looked at the region that Ice-Cube said the neu-

"Everything points to this as the ultra-bright, energetic source — a gorgeous source."

trino came from, we noticed that this blazar had been flaring more than ever before," says Regina Caputo, an astrophysicist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, who is Fermi-LAT's analysis coordinator

On 28 September, the Fermi-LAT team sent out an alert to reveal this finding. It was at that point that other astronomers got excited. Ice-Cube has detected about a dozen such high-energy neutrinos each year since it started operating in 2010, but none had been associated with a particular source in the sky. "That's

what made the hair stand at the back of the neck," Fox says.

Researchers with IceCube and Fermi-LAT calculated the odds that the flare and the neutrino were related, rather than coming from the same direction in the sky by chance. They found that likelihood to be good, although not at the level of statistical significance required for claiming a discovery in physics^{1,2}.

A major missing piece of information was the blazar's distance from Earth, says Simona Paiano of the Astronomical Observatory of Padua in Italy. To measure it, she and her team booked 15 hours of observing time on the world's largest optical telescope, the 10.4-metre Gran Telescopio Canarias on La Palma, one of Spain's Canary Islands. They found the blazar to be around 1.15 billion parsecs (3.78 billion light years) away³.

Together, the data pinpoint the likely source, says Kyle Cranmer, a particle physics and dataanalysis expert at New York University, but "the observation isn't unambiguous", he cautions. "More follow-up is needed to conclusively establish blazars as a source of high-energy neutrinos."

- 1. IceCube Collaboration. *Science* **361**, 147–151
- 2. IceCube Collaboration et al. Science **361**, eaat1378 (2018)
- 3. Paiano, S., Falomo, R., Treves, A. & Scarpa, R. *Astrophys. J. Lett.* **854**, L32 (2018).

EPIDEMIOLOGY

Chinese cities scan sewers for signs of illegal drug use

Privacy concerns and cultural differences could limit the technique's use in other nations.

BY DAVID CYRANOSKI

ozens of cities across China are applying an unusual forensic technique to monitor illegal drug use: chemically analysing sewage for traces of drugs, or their telltale metabolites, excreted in urine.

One southern city, Zhongshan, a drug hotspot, is also monitoring waste water to evaluate the effectiveness of its drug-reduction programmes, says Li Xiqing, an environmental chemist at Peking University in Beijing who is working with police in these cities.

Li says Zhongshan police have already used the technique to help track down and arrest a drug manufacturer. He says a handful of cities are planning to use data from waste water to set targets for police arrests of drug users, some as early as next year.

Although illegal drug use has been monitored through wastewater-based epidemiology (WBE) in other countries, including Belgium, the Netherlands, Spain and Germany, most studies have collected data for epidemiological research rather than for setting policies. "The noteworthy part is that China seems to be actually acting on the technique," says Daniel Burgard, a chemist at the University of Puget Sound in Tacoma, Washington.

Last month, Chinese President Xi Jinping said that the country's war on drugs was tied to national security and the welfare of the Chinese people. Li says the central and local governments will invest at least 10 million yuan (US\$1.5 million) in WBE monitoring by the end of the year. He expects the figure to at least double annually for the next few years.

Li is pushing for the method to be used

internationally, including as part of the United Nations' drug control policies. "The experience and lessons from the application of WBE and its adoption by the Chinese drug police in their daily management will be very relevant for other countries," he argues.

But many issues, such as how police should be allowed to analyse the data, the need for safeguards to prevent the data from being misused, and privacy concerns, need to be ironed out. Some researchers are sceptical that the method will be adopted successfully in other countries.

DRUG USE

To show that WBE reflects drug use in the community, a number of studies have compared drug levels detected in sewage with other data sources on drug use, such as the amount





Chinese officers destroy seized drugs.

of drugs seized by police, and user surveys. A 2016 study in eight European cities found a strong correlation between the amount of cocaine detected in waste water and data from drug seizures (J. A. Baz-Lomba et al. BMC Public Health 16, 1035; 2016). However, in the case of metamphetamines, the correlation was not as strong.

Researchers around the world generally agree that WBE can reliably estimate drug use, says Shane Neilson, the head of Determination for High Risk and Emerging Drugs at the Australian Criminal Intelligence Commission in Canberra. "The science and findings are

globally consistent and comparable," he says. The technique is also used by health researchers to detect other substances excreted by humans, such as signs of bacteria and viruses.

Zhang Lei, an environmental policy researcher at Renmin University in Beijing who collaborates with Li, notes that WBE studies are a more objective way of measuring whether government initiatives to reduce drug use in the community are working. She says that relying solely on conventional methods for monitoring changes in drug use, such as the number of arrests of users or the number of drugs being seized by police, can be misleading

because they are indirect measures. "WBE offers an unequivocal measure of the effectiveness of efforts," says Zhang.

Li and his team put this to the test when they measured two popular synthetic drugs, methamphetamine and ketamine, in waste water across China two years after local and national agencies launched campaigns to crack down on drug use and manufacturing in 2013. Zhang's team found that after these initiatives, methamphetamine use dropped by 42% and ketamine use decreased by 67%. Li thinks the drop in drug use is a result of police campaigns.

OTHER COUNTRIES

Jose Antonio Baz-Lomba, a researcher at the Norwegian Institute for Water Research in Oslo, says the growing evidence that the technology is a reliable measure of drug use should encourage other international police authorities to take WBE seriously and start collaborating with researchers.

But Carsten Prasse, an environmental-health researcher at Johns Hopkins University in Baltimore, Maryland, argues that cultural and political differences between countries will have a substantial effect on this research. "In China, the general population is used to following the directions given by the government, and privacy-related issues don't seem to be a major concern — the situation is totally different in the United States," he says.

Prasse says the potential implementation of wastewater-based drug monitoring needs to be discussed in the community, not only between scientists and law enforcement. "WBE represents a powerful new tool to assess drug consumption in our cities, but there is still a lot of work to do before it can be implemented on a larger scale," he says.

PUBLISHING

Experimental openaccess deal ends

Science's pilot contract with the Gates Foundation aimed to solve a policy conundrum that affects several journals.

BY RICHARD VAN NOORDEN

he publisher of *Science* last month ended a pilot partnership that allowed openaccess (OA) publishing for researchers funded by the Bill & Melinda Gates Foundation.

The trial was an effort to accommodate a policy clash between the Gates Foundation, which has enforced strict OA demands since

2017, and publishers running subscription journals that don't comply with those terms. So far, 26 papers have been published in *Science* and 4 sister subscription journals as part of the 18-month experiment, and more might appear, says a spokesperson for *Science*'s publisher, the American Association for the Advancement of Science (AAAS) in Washington DC. Neither the Gates Foundation nor the AAAS commented on why the deal ended.

Under the contract, the Gates Foundation paid the AAAS a lump sum of around US\$100,000 for a trial first year, during which 16 papers appeared. The two organizations then extended their partnership for another six months, and continued their contract on "similar terms", but have agreed to keep the extra amount paid confidential, says Bryan Callahan, an external-relations officer at the Gates Foundation.

Meanwhile, two other influential journals, The New England Journal of Medicine (NEJM) and Proceedings of the National Academy of Sciences (PNAS), quietly changed their policies last year to offer a permanent OA publishing route for Gates grant holders. And although Nature has not made a specific agreement with funders, it has published some papers under OA terms, including two Gates-funded papers this year. (Nature's news team is editorially independent of its journal team and of its publisher, Springer Nature.)

The Gates Foundation, based in Seattle,

▶ Washington, is a global health charity that spent \$4.6 billion in 2016, much of it allocated to research. Each year, more than 2,000 papers are published from projects it funds. The foundation stipulates that these papers, and their data, must be made open.

It's not the only research funder to have such rules, but its policy is stricter than most, because it demands that papers are made free to read immediately on publication, rather than permitting a six-month delay as some subscription journals require. And the papers must not only be free to read, but also be posted under a 'CC-BY' licence that allows their contents to be reused without restrictions, for example through republication, even for commercial purposes. When the Gates policy came into force at the beginning of 2017, it clashed with the rules of subscription journals including Nature, Science, NEJM and PNAS, meaning that researchers could not publish Gates-funded work in these journals.

In February that year, however, the AAAS and Gates announced their partnership. On 1 March, *NEJM* changed its own policy. The medical journal generally makes articles free to read on its website six months after publication, but it agreed to make Gates-funded

articles free to read immediately, says Jennifer Zeiss, communications and media-relations manager for the NEJM Group. It also agreed to simultaneously make available a CC-BY licensed 'author final version' of the paper, which includes revisions made after peer review but lacks final NEJM editing. These appear online in the PubMed Central database.

"At present time, NEJM does not have this arrangement with other funders," Zeiss says.

And in September 2017, PNAS — which also already makes

"If Gates had refused to pay the prestige tax, it would not have lost grant applications."

papers free to read on its site six months after publication — began offering an OA option under a restrictive licence that does not permit commercial reuse or republication. The journal also decided to offer a liberal CC-BY licence for authors whose funders mandate it, a spokesperson says.

Nature does not have a specific OA policy for Gates grant holders, but the issue is still under discussion, and the journal does occasionally publish papers, which can include those with Gates funding, under a CC licence, says a spokesperson for Nature Research, the portfolio of journals that includes *Nature*. The journal has published more than 30 CC-BY OA papers since 2017, according to an analysis by *Nature*'s news team, including the two by Gates-funded researchers.

Peter Suber, director of the Harvard Open Access Project and the Harvard Office for Scholarly Communication in Cambridge, Massachusetts, characterizes the AAAS pilot as a compromise whereby Gates paid the publisher a "prestige tax" for the specific OA terms it wanted.

"To me, the deal was unnecessary and undesirable. A wide range of high-quality journals were already compatible with the Gates publishing terms. If Gates had refused to pay the AAAS prestige tax, it would not have lost grant applications from first-rate researchers," Suber says. "I'm glad to see it come to an end."

Other funders haven't imposed terms as stringent as Gates's, notes Stephen Curry, a structural biologist at Imperial College London, but he praises the stance. "Gates are right to stipulate immediate OA as a condition of funding, especially in an area of such importance to global public health."

ASTRONOMY

Ten new moons spotted orbiting Jupiter

Planet now has 79 known satellites, including one on a collision course with its neighbours.

BY ALEXANDRA WITZE

A stronomers have discovered 10 small moons orbiting Jupiter, bringing its total to 79 — by far the most moons known around any planet. One of the finds is an oddball that moves in the opposite direction from its neighbours.

Together, the moons help to illuminate the Solar System's early history. The existence of so many small satellites suggests

that they arose from cosmic collisions after Jupiter itself formed, more than 4 billion years ago.

"They did not form with the planet, but were likely captured by the planet during or just after the planet-formation epoch," says Scott Sheppard, an astronomer at the Carnegie Institution for Science in Washington DC. He and his colleagues announced the discovery on 17 July.

Sheppard's team typically hunts for objects

in the very distant Solar System, out beyond Pluto, and sometimes spots planetary moons during these searches. Last year, the group reported two additional Jovian moons. In this case, the scientists were looking for a putative unseen massive planet popularly known as Planet Nine. Jupiter was in the same part of the sky, so they were able to hunt for moons as well.

To discover new Solar System bodies and calculate their orbits, the researchers

MORE ONLINE

TOP NEWS



CRISPR gene editing produces unwanted DNA deletions go.nature. com/2lkzfmq

MORE NEWS

• South Africa celebrates the completion of a gigantic, supersensitive radio telescope go.nature.

• Tools from China are the oldest hint of human lineage outside of Africa go.nature.com/2nzrgs2

NATURE PODCAST



DNA nanostructures; climate-altering bacteria; and a chemical reactionfinding robot nature. com/nature/podcast WELLCOME/CC BY NC

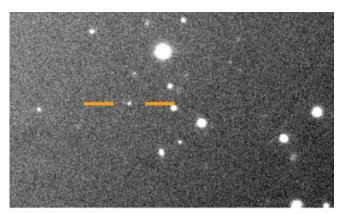
photograph the same part of the sky weeks or months apart. They then look for objects that shift position between the two images, relative to the background stars. The team first spotted most of the new Jovian moons using the Blanco 4-metre telescope at the Cerro Tololo Inter-American Observatory in Chile, and followed up with further observations at other telescopes.

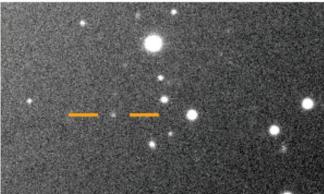
STRANGE SATELLITES

All the newfound moons are small, between about 1 and 3 kilometres across. Seven of them travel in remote orbits more than 20 million kilometres away from Jupiter, and in the opposite direction from the planet's rotation. That puts them in the category known as retrograde moons.

The eighth moon stands out because it travels in the same region of space as the retrograde moons, but in the opposite direction (that is, in the same direction as Jupiter's spin). Its orbit is also tilted with respect to those of the retrograde moons. That means

it could easily smash into the retrograde moons, pulverizing itself into oblivion. It may be the leftovers of a bigger cosmic collision of this nature in the past, Sheppard says.





These images show the movement of the Jovian moon dubbed Valetudo (labelled in yellow) relative to the background stars.

Jupiter's moons are named after gods with connections to the mythological deities Jupiter or Zeus. Sheppard has proposed naming the oddball Valetudo, after one of Jupiter's descendants, the Roman goddess of hygiene and health

The ninth and tenth newfound moons orbit closer to Jupiter, moving in the same direction as the planet.

Had all these small moons formed at the same time as Jupiter, they probably would have been captured by the gas and dust still swirling around the newborn planet, and have been engulfed. Their existence suggests that they are leftovers of later collisions between space rocks that left the debris encircling Jupiter.

If astronomers can work out the history of these collisions, they could also determine the sizes of any satellites that were pulled into the orbit of a young Jupiter. "That's the big question, and that's what makes these ten new moons interesting," says Douglas Hamilton, an astronomer at the University of Maryland in College Park. "How can we link all this to how planets formed?"

Sheppard says there might still be a few more moons of Jupiter to discover — as yet unseen because

they were hiding in the Sun's glare when the scientists were looking. Saturn, the runner-up to Jupiter in the moon competition, has 62 known satellites. ■

MEDICAL RESEARCH

Gene therapy in mouse fetuses treats deadly disease

 $The \,method \,could \,minimize \,damage \,from \,disease \,if \,a \,condition \,is \,diagnosed \,in \,utero.$

BY HEIDI LEDFORD

ene therapy administered in the womb could be used to treat a deadly genetic disease, a study in fetal mice suggests. The results could add to the increasingly popular approach of using prenatal gene therapy to minimize the damage wrought by some genetic diseases. The US Food and Drug Administration approved the first gene therapy for adults and children last year, and more treatments are crowding pharmaceutical pipelines around the world.

Simon Waddington, the lead author of the latest study, says he used to meet with shocked stares when he talked about treating fetuses

with gene therapy. "It had gotten to the point where I'd given up on telling people that fetal gene therapy is a good idea," says Waddington, who studies gene therapy at University College London. "And now, not infrequently, people turn to me and say, 'You know what would be a good idea? Fetal gene therapy."

The mouse study, published on 16 July in *Nature Medicine*¹, uses prenatal gene therapy to tackle a condition — acute neuronopathic Gaucher's disease — caused by mutations in a gene called *GBA*. These mutations disrupt the breakdown of a particular fatty molecule, or lipid. As a result, the lipid accumulates in brain cells and other parts of the body, contributing to organ dysfunction.

The study looks at whether the disease can be treated by using a virus to supply normal copies of *GBA* to a developing fetus. That could minimize the irreparable brain damage that arises as the lipid accumulates.

Some forms of Gaucher's disease can be treated by supplying normal copies of the GBA enzyme to break down lipids, but that enzyme cannot cross from the blood into the brain. Children with acute neuronopathic Gaucher's disease rarely live past two years.

DIFFICULT CROSSING

The condition is so devastating that colleagues were sceptical about his team's ability to treat it, says Waddington. "People told me,



'You're not going to touch this."

One hurdle was simply getting the virus to carry the healthy gene into the brain. Viruses used in previous tests had to be injected directly into the brain, and then they diffused only a short distance from the injection site. But in 2009, researchers showed^{2,3} that a particular virus, simply injected into the blood, could reach the central nervous system. From there, it dispersed throughout the brain.

Waddington began working with mice, loading up the virus with a normal copy of *GBA* and looking for ways to express it specifically in the central nervous system. His team tested its virus in fetal mice carrying *GBA* mutations that cause symptoms similar to neuronopathic Gaucher's disease. Such mice normally live for only 15 days after birth; treated mice, however, survived for at least 18 weeks and were able to move about normally.

FETAL FRONTIER

The work is impressive, says Tippi MacKenzie, a fetal-medicine specialist at the University of California, San Francisco. MacKenzie has been conducting a clinical trial of prenatal stem-cell transplants. "Fetal gene therapy or enzyme-replacement therapy may be the next frontier," she says. "It is wonderful to see this kind of rigorous research, to take us one step further."

Treating fetuses has several potential advantages. Chief among them is the potential to minimize the damage caused by a genetic disease. Some of these conditions — such as neuronopathic Gaucher's disease and spinal muscular atrophy — can cause irreversible symptoms before birth.

It is also easier to administer some therapies to the brain in a developing fetus than

"People told me, 'You're not going to touch this.'" in an adult or child, because the bloodbrain barrier — a membrane that prevents some molecules from crossing into

the brain from the blood — is more permeable.

"Even one day after birth, it's harder to get into the brain," says Jerry Chan, an obstetrician and gynaecologist at Duke-NUS Medical School in Singapore.

And the fetal immune system is also still developing, making it less likely to recognize the newly expressed protein as foreign. Adult immune systems sometimes generate antibodies against the new protein, which can prevent it from carrying out its function.

WEIGHING THE RISKS

Chan and others have previously tested fetal gene therapy to treat haemophilia in mice and macaques, and Chan expects that there will be interest in doing so for several metabolic diseases similar to Gaucher's.

But there are risks. Researchers developing a prenatal gene therapy must think not only about the fetus, but also about the mother, who will inevitably receive a dose of treatment as well, says Chan.

And clinicians have to be absolutely certain that the mutation they've found will cause disease, notes Waddington. This may mean combining genetic tests with other tests performed *in utero*, to confirm the disorder.

"We're now at the point where it's possible to diagnose these diseases," he says. "It's making people think: maybe we should be doing this." ■

- Massaro, G. et al. Nature Med. https://doi. org/10.1038/s41591-018-0106-7 (2018).
- Foust, K. D. et al. Nature Biotechnol. 27, 59–65 (2009).
- 3. Duque, S. et al. Mol. Ther. 17, 1187-1196 (2009).

CORRECTION

The News Feature 'Bat man' (*Nature* **559**, 165–168; 2018) talked of the Egyptian fruit bats under study being collected in the Jordanian hills instead of the Judean hills. It also incorrectly labelled the picture of a neural logger as being a GPS logger.

SCIENCE UNDER SIEGE

Uncertainty, hostility and irrelevance are part of daily life for scientists at the US Environmental Protection Agency.

BY JEFF TOLLEFSON

he day Donald Trump took office as US president, the mood was sombre at the main research campus of the Environmental Protection Agency (EPA) in Durham, North Carolina. As scientists arrived for work, they saw pictures of former president Barack Obama and the previous EPA administrator, Gina McCarthy, coming down off the walls. Researchers had reason to be anxious: Trump had threatened many times during his campaign to shutter the EPA, and he had already taken steps along that path. Weeks before he moved into the White House, Trump had nominated Scott Pruitt to head the agency — a man who had spent his career filing lawsuits to block a variety of EPA regulations.

When Trump put his hand on the Bible to take the oath of office on 20 January 2017, many EPA scientists kept their heads down. They wondered who might be fired first, and they warned each other to censor their e-mails, for fear that the new administration would monitor communications for any comments criticizing it.

Dan Costa wasn't so worried. After nearly 32 years working at the EPA, he had seen the agency weather many political storms, and he had not lost sleep over the prospect of working for Pruitt and Trump. When inauguration day came, Costa streamed Trump's speech on his computer and went straight back to work.

"There was a lot of fear and anticipation, but I figured we would push through it," says Costa, who at the time headed the department's air, climate and energy research programme.

Over the next 18 months, however, Costa would grow increasingly concerned about the Trump administration's impact on the agency. Since assuming power, this administration has launched more assaults on the EPA than on any other science agency. The president has sought to slash its budget by nearly one-third, and Pruitt's team has tried to

weaken the part that science plays in setting environmental regulations. He barred some top researchers from participating in EPA advisory panels, and replaced them with scientists who are more friendly to industry. All of this has elevated the power of corporations to influence the rules that govern chemicals and pollutants.

But what is it like for the more than 1,000 scientists working at the EPA itself? To find out, *Nature* has conducted dozens of interviews over the past year and a half with current and former agency staffers.

The interviews show that day-to-day work has changed little for many EPA researchers. They continue their investigations into everything from ecology and toxicology to hydrology and air quality, in an effort to bolster the scientific foundations for health and environmental regulations.

What has damaged researchers' morale is the endless uncertainty about all aspects of their work, and the thinly veiled hostility from the administration. It's the onslaught of media stories about budget cuts, staff lay-offs and efforts to weaken environmental and health regulations. It's the ever-growing scent of scandal as Pruitt came under media fire for lavish spending with government funds, allegedly using his office to find a lucrative job for his wife, among other potential ethical breaches. Pruitt denied any wrongdoing, but ultimately resigned on 5 July.

What most troubles many EPA scientists is the Trump administration's systematic and unprecedented effort to undermine the way in which science is used by the agency. Scientists there say they and their work have been largely ignored by senior EPA leadership. And despite Pruitt's resignation, few expect the administration's overarching EPA strategy to change once Trump appoints a new administrator. For now, the leadership reins fall to Andrew Wheeler, a former coal lobbyist. In a pair of tweets announcing Pruitt's resignation, Trump said that Wheeler would "continue on with our great and lasting EPA agenda".



Many researchers say that this strategy could subvert the scientific process altogether and put tens of thousands of lives at risk each year, as a result of weakened regulations on pollutants and potentially hazardous chemicals.

The turmoil has affected everyone. Most have kept their heads down, hoping that science will somehow prevail. Many have censored their own language, shunning words such as 'climate' or 'global warming' to avoid attention. Some have delayed retirement to keep the agency functioning. Others have quit.

"There's a lot of fear, a lot of angst and anxiety, and employees don't know what to do," says Kyla Bennett, director of science policy at the environmental organization Public Employees for Environmental Responsibility (PEER) in North Easton, Massachusetts. PEER works directly with many government whistle-blowers. "This is unlike anything we've ever seen," Bennett says.

Costa has watched the situation deteriorate. As he tried to carry on his own work, his mood grew darker and more philosophical. Eventually, he realized he had to leave. "They are acting with such impunity, and with no accountability," he says of the administration. "It's just unfortunate, and scary."

THE FIRST 100 DAYS

At the beginning of Trump's presidency, Costa's long history with the agency helped him to cope. A toxicologist by training, he joined the EPA in 1985 under president Ronald Reagan, looking at the physiological effects of pollutants. He arrived shortly after the tenure of Anne Gorsuch, a staunchly conservative administrator — much like Pruitt — who had slashed budgets and weakened environmental protections during her time heading the EPA from 1981 to 1983. Yet Costa watched the agency

"THERE'S A LOT OF FEAR, A LOT OF ANGST AND ANXIETY, AND EMPLOYEES DON'T KNOW WHAT TO DO."

slowly bounce back.

That episode served as a reminder that the institution is larger than any individual, Costa told *Nature* in early March 2017, during one of a series of interviews initially conducted off the record because he didn't have permission to talk to the press. He later agreed to bring the entire series on the record.

At the time, stories were swirling in the media about

censorship and looming budget cuts. Costa said that much of it was probably true, but he also stated that such stories can grow out of proportion. "It's not like there are memos coming down. It's just rumours," he said about talk of censorship. "And in the absence of good information, it's easy for people to create their own demons." Younger scientists had been coming to him for advice, asking whether they should start looking for jobs, and his advice was simple: don't panic.

The Trump administration soon made its intentions clear. On 16 March 2017, it released a proposal to slash the EPA's US\$8.2-billion budget by 31% and eliminate some 3,200 of the agency's 15,000 positions.

Among the hardest hit in the budget proposal was the division where Costa and some 1,100 other scientists worked: the Office of Research and Development (ORD). As the main science arm of the agency, the ORD has helped to lay the technical foundation for modern environmental regulation in the United States. The Trump administration



had proposed nearly halving its budget from, \$483 million to \$250 million, which left scientists there stunned.

"Management at all levels are trying to reassure employees, but you can't help but worry," Lesley Mills told *Nature* at the time. "These are people who are dedicated to public service, and they feel like they are being treated as an enemy," said Mills, an EPA biologist in Narragansett, Rhode Island, and a union representative.

She and others knew that the planned cuts might never happen. Congress has authority over the budget in the United States and often decides to override presidents' budget requests. And legislators — including many important Republicans — were unusually sceptical about Trump's first proposal. Ignoring the administration's calls for sharp cuts to EPA, on 30 April the Republican-controlled Congress approved a relatively mild reduction of 1% for the remainder of the 2017 fiscal year.

It felt like a triumph for many scientists, but Costa was already beginning to change his tone. When he attended an inaugural March for Science event in April in Raleigh, two EPA scientists with him instinctively ducked and threw their hands up to hide their faces when a news photographer approached. They told him that they didn't want to encounter questions later from political leadership at the agency.

Costa also found himself encouraging one promising young postdoc to apply for a position elsewhere, because he thought EPA jobs were unlikely to open up in the next few years. He knew of managers who had told younger scientists to take the word 'climate' out of document headlines. "That sends all sorts of ripples through the organization," he said in May 2017.

At the same time, Costa was making his own changes. He was quietly trying to expand the air, climate and energy research programme that he ran to advance a new line of science, protect his team and avoid attention from higher authorities at the agency. As he sat in meetings and drafted reports, he talked increasingly about public health and wildfire smoke rather than just the industrial air pollutants that his programme had historically focused on. Costa described the proposed shift in scientific focus as a positive change that would define a useful agenda for his programme without limiting the science that it could pursue, in part because climate change, air quality and public health are all interrelated.

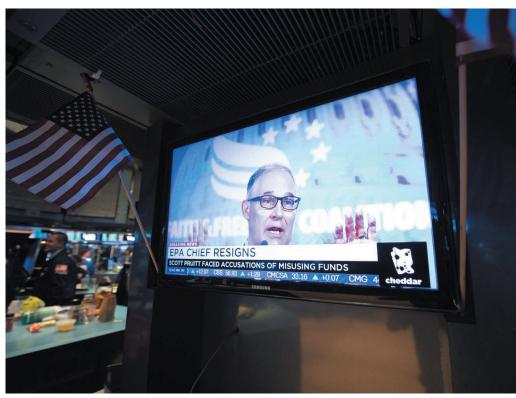
"I don't want to sit back and wait" for any restrictions to be imposed by political leaders at the agency, Costa said. "I want to occupy the space before they do, because they are essentially clueless."

A GROWING RIFT: SUMMER 2017

All the while, Pruitt was busy trying to roll back environmental regulations put in place by Obama — including regulations that Pruitt had challenged while serving as Oklahoma attorney-general. On 28 March, Trump authorized Pruitt to repeal landmark regulations intended to curb greenhouse-gas emissions from existing power plants. The next day, Pruitt declined to ban a powerful pesticide called chlorpyrifos, overruling agency scientists who had previously determined that the chemical had negative impacts on brain development in children (see go.nature.com/2n7pofa).

What alarmed scientists about these and other actions was not so much that Pruitt and Trump were moving in a different political direction from the Obama administration; government scientists are used to that. But under previous administrations, regardless of political stripe, there was at least some deference paid to scientists.

That all changed with Trump. Pruitt and his senior political



EPA administrator Scott Pruitt resigned on 5 July.

appointees — often dubbed the "politicals" — rarely consult with career scientists. In many cases, scientists were left dumbfounded, in part because the complete lack of consultation with agency experts could end up hurting Pruitt's own agenda. By bypassing EPA scientists and ignoring their findings, his team ran the risk of weakening the EPA's defence in the many lawsuits that states and environmental groups were filing against the agency.

"The politicals literally do not talk to the career people," says one senior scientist. That researcher and nearly all active EPA staff interviewed for this story sought anonymity because they were not authorized to talk to the press. "They just do what they want, and then they inform us," says the senior researcher.

In an effort to cope with the new reality, another senior official said, career scientists looked for areas of common ground with the leadership and, in a curious dance, both sides tiptoed around the issue of climate change. "It's like Voldemort — he who shall not be named," the official said in mid-2017.

"There are weeks when everyone in the office is just chugging along like normal," says one mid-level scientist. Inevitably a scandal arises, he continued, "and then for a day or two you feel like you are in a fog".

Although they carry on with their work, many scientists feel as if their efforts don't matter to the top of the agency. Within the Office of Research and Development, exchanges with senior EPA leadership nearly always go through an intermediary: Richard Yamada. Yamada, deputy assistant administrator of the office, was willing to communicate ideas up the chain, according to multiple scientists, but he often seemed adrift on technical or scientific issues.

Yamada asked such odd questions during one video conference that researchers in the meeting found themselves looking at each other in confusion. "You go into these briefings, and you have no idea what the questions are going to be." (The EPA did not grant *Nature*'s request to speak to Yamada and has not responded to multiple requests for comment on the allegations in this article.)

The rift between the scientists and EPA leadership was fully exposed in late July 2017, when news broke that Pruitt's team was circulating a list of names of climate sceptics. Many assumed the EPA was looking for sceptics to participate in a proposed debate about the validity of

climate science or, potentially, for appointments to science-advisory positions. The proposal came as the EPA was conducting a technical review of a government assessment of current climate science. People from both inside and outside the agency had raised concerns about whether Pruitt — who as recently as four months earlier had questioned the scientific consensus on climate change — and his political appointees would meddle with the document.

Pruitt's team eventually let the scientific assessment move forward. Costa and others gave the agency credit for that decision. "They have the authority to slow these assessments down or stop them, if they want," he said at the time. "In spite of all of the rhetoric, it's going through a reasonably normal process."

For Costa, it was evidence that in many senses, the EPA's leadership doesn't really care about what scientists do — unless and until it gets in the way of Trump's agenda to roll back regulations on industry. But as it turned out, the administration was just getting started.

TENSIONS GROW: AUTUMN 2017

On 31 October — Halloween, no less — Pruitt dropped a bombshell on the scientific community in the United States. He announced that scientists with active EPA grants would be banned from serving on the agency's main science advisory board (SAB) or on a separate committee focused on air regulations. Such committees provide peer review of the science underlying most EPA regulations; Pruitt's decision prevents some of the nation's top environmental scientists from taking part in that process.

Pruitt justified his action with a damning charge: research grants provided by the EPA, he said, could bias scientists and the advice they give to the agency. Scientists were shocked because this policy stands in sharp contrast to those of other science agencies, such as the National Institutes of Health, and also because researchers with industry support were not similarly barred from EPA advisory boards. The surprises didn't end there. Pruitt also called for limiting the tenure of board members, which would force even more scientists to cycle off the board. Pruitt would thus get to select replacements more quickly.

As a result, 18 of the 44 members of the science advisory board are now Pruitt appointees. By the end of September, Trump's team at the EPA will have appointed roughly two-thirds of the council, says Christopher Zarba, who until his retirement in February managed the board's activities at the EPA. Many fear the board will increasingly hew to the desires of powerful interests involved in everything from chemicals to energy and manufacturing.

Perhaps most significantly, Pruitt selected Michael Honeycutt to chair the SAB. Honeycutt is a toxicologist with the Texas Commission on Environmental Quality in Austin, Texas, who has long opposed stricter air-quality standards. (Honeycutt told *Nature* that he hopes he will be judged on the basis of the job he does as the chair of the board.) And Pruitt appointed Tony Cox, an industry-friendly consultant who has challenged scientific studies linking air pollution and human mortality, to lead the Clean Air Science Advisory Committee (CASAC). By statute, that group must review the science before the agency updates its core air-quality standards.

By the time of those appointments, Costa was already growing weary of the attacks on science, but he still saw room to do some good by reorienting his programme. Costa had long lobbied to focus more research on wildfires because they contribute a large fraction of the fine-particulate pollution across the country, he says. The agency, however, had devoted its resources for decades to tackling industrial air pollution.

With Trump and Pruitt in office, Costa thought the time was right to give his programme a new mission by including a focus on wildfires. In early December, the air-quality regulatory division — the primary customer for the air-research programme — informally endorsed Costa's new research agenda. With that small victory, Costa, who was 69 at the time, decided it was time to leave.

"I certainly didn't want to be a rat jumping ship," he said. But with five children, five grandchildren, a new riding lawnmower and a sudden dedication to science activism, Costa has more than enough to keep himself busy on the outside. "I just didn't think I would do well in the current atmosphere."

On 5 January, two weeks before Trump celebrated his first year in the White House, Costa went into the EPA one last time. His co-workers had already thrown him a party — complete with a Beatles-themed musical skit. As the end of his final day, Costa packed up the remaining boxes, turned in his parking pass and headed home.

PRUITT RESIGNS: SPRING 2018

Over the ensuing months, more news emerged about Pruitt's alleged ethical transgressions. There were investigations, congressional hearings and endless speculation about how long the embattled administrator could retain the favour of his mercurial White House boss. In the end, Pruitt would stay on for another six months — and drop yet another bomb on scientists at the agency.

On 24 April, Pruitt announced a proposal that would prevent the EPA from using any research in its regulatory decisions unless the underlying data and methods are publicly available. He did so in the name of transparency, but scientists and other experts immediately fought back.

The problem, they said, is that privacy restrictions — such as ones governing medical records — often limit the data that can be released from epidemiological studies, to protect patients' identities. Pruitt's proposal could therefore eliminate much of the core epidemiological research that the EPA has used to help justify air-quality regulations. It was, in their view, just another effort to prevent the agency from developing meaningful health and environmental regulations. In one analysis released in April, a group of former EPA officials found that Pruitt's policy, if implemented two decades ago, could have precluded regulations that now prevent some 50,000 deaths each year from air pollution (see go.nature.com/2zmrmgt).

When the news broke, Costa was so incensed that he reached out to *Nature* from retirement. "Keep your eyes on this: it's an IED [improvised

"IT'S LIKE VOLDEMORT — HE WHO SHALL Not be named."

explosive device] designed and set to destroy the agency's ability to do its job," Costa wrote in a text message. Pruitt, he continued, "is a slick bastard".

A day after the rule was announced, a poster of Pruitt signing the rule, with grand proclamations about transparency in science, appeared at the entrance of the ORD's main

building in central Washington DC. For many scientists, it was yet another insult.

"That poster said, 'I've got you, and there's not a damn thing you can do about it," says the senior scientist at the EPA. "They are making sure that we understand that there's a new sheriff in town."

For his part, Costa says he doesn't have any regrets. He is enjoying the summer in a remote stretch of coastal Rhode Island, where he used to spend time during his youth. But clearly he hasn't let go — in part, perhaps, because he still doesn't know how the story will end. "The light at the end of the tunnel just doesn't seem to be there," he said in late May.

When the news of Pruitt's departure came down on 5 July, Costa was dawdling in the garage. His wife ran out of the house to tell him and his mobile phone lit up with texts from friends, family and former colleagues at the EPA. Costa was relieved, if not surprised. Looking forward, he hopes that Wheeler — who spent four years at the agency in the early 1990s — will not be so quick to ignore science and scientists, even if he does toe the Trump line.

And after a few recent conversations with former staff members, Costa seems newly encouraged that they will keep the embers burning until the political winds shift again and sweep away Trump's team. "In some senses, I think of it like the locusts," he says. "They come, they wipe out the crops and then they leave."

Jeff Tollefson reports for Nature from New York.

COMMENT

MACHINE LEARNING Three ways researchers must make algorithms more fair p.324

SUMMER BOOKS Survey of health inequities in African Americans **p.328**



IRAN Trump will not crush scientists who have withstood decades of sanctions **8.331**



China is clamping down on unauthorized coal-fired factories, such as this one, to reduce carbon emissions.

Beat protectionism and emissions at a stroke

Applying carbon charges, not trade tariffs, to imports would bolster the Paris Agreement, argue **Michael Mehling** and colleagues.

wo huge multilateral issues — free trade and climate change — top policymakers' agendas in 2018. This offers a chance to couple them.

More and more countries are shielding domestic producers from foreign competition

— a process known as protectionism. Since January, US President Donald Trump has slapped tariffs of up to 50% on many imports, including washing machines, solar cells, soya beans, steel and aluminium. Hopes that allied countries would be exempt were dashed

after a tumultuous G7 meeting in June.

Economies affected have begun to respond in kind. China hit back with levies on US\$34 billion worth of US goods. The European Union increased tariffs on jeans, motorbikes and bourbon imported from the United ▶ States. And Trump has since threatened to add tariffs on another \$200 billion worth of Chinese goods. A trade war is unfolding.

Meanwhile, nations are reviewing the pledges they made to cut emissions as part of the 2015 Paris Agreement. Everyone knows that current pledges will not keep global warming below the 'safe' limit of 2 °C above preindustrial levels — even if all nations deliver on their promises. The question is how to strengthen actions so that emissions drop sharply once the Paris framework takes effect in 2020.

The Paris process has two main problems. First, the pledges are uneven. Countries that do little will benefit from hefty cuts made by others. In the ultimate free ride, the United States will withdraw from the Paris Agreement in 2020, leaving others to do more. Second, carbon emissions 'leak' across borders. A country can keep its budget low by buying carbon-intensive goods made elsewhere. Some regions, such as Western and Northern Europe, import a considerable share of high-emission goods, allowing them to emit less themselves (see 'Carbon balance').

Over the next two years, there will be a flurry of activities relating to trade and climate change. This is a perfect opportunity to tie together the two agendas.

Governments should levy a carbon charge on imports. These 'border carbon adjustments' (BCAs) would level the emissions playing field by imposing the same economic burden on domestic and external manufacturers. Producers would lose the incentive to manufacture goods in places with weaker carbon regulations. Trade partners would then prefer to manufacture and export low-carbon products to avoid penalties.

Political interest in BCAs is growing. In 2017, French President Emmanuel Macron called them 'indispensable' for European climate leadership, and Canadian environment minister Catherine McKenna recommended closer scrutiny. Mexico included them in its Paris pledge. But there are fears of retaliation and some confusion over the legality of BCAs. Here's how nations could proceed.

CLIMATE-SMART TRADE

BCAs should be applied to imported goods in line with their carbon footprint (the total of all the emissions released during their manufacture). A country might impose a fee on imported steel if its carbon footprint is higher than that of domestic steel, for example. Alternatively, governments might require importers to purchase emission allowances in carbon markets. The

US House of Representatives backed such an approach when it passed the Waxman–Markey Bill in 2009; however, the bill failed to reach a vote in the Senate.

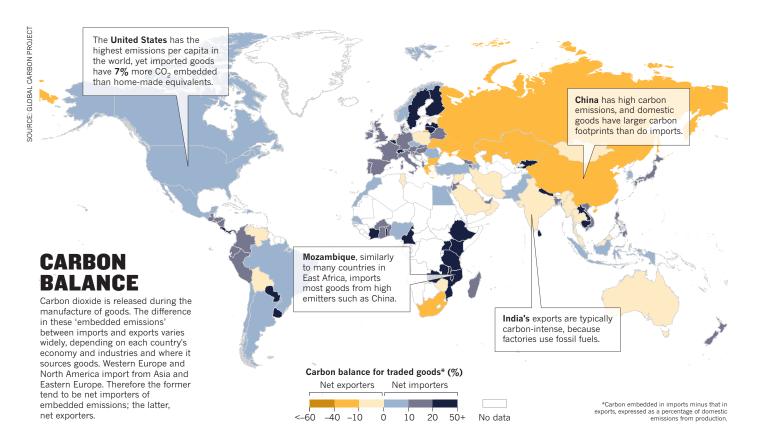
California is the only jurisdiction to have introduced BCAs, in its energy market. Since 2013, electricity delivered into California from neighbouring states has been subject to the same carbon constraints as that generated domestically. This has stopped electricity suppliers from shifting power generation to states with lax climate policies.

So far, efforts to limit emissions leakage have been less efficient. Some countries, including Germany, offer regulatory relief or compensation payments to domestic emitters to persuade them not to relocate. But such measures have undermined their other climate policies. For example, when the EU handed out emissions-trading permits for free, it weakened incentives to curb emissions and produced windfall profits for some energy-hungry companies.

BCAs, by contrast, bolster climate policies. By restricting trade in carbon-intensive goods, they accelerate decarbonization even in countries with weak regulation. They also appeal to policymakers, manufacturers, trade associations and



DD SPOTH/NYT/REDUX/FYEV



labour unions who are concerned about a nation's economy and jobs. Imposing BCAs on imports from the United States would prove politically popular with these groups. It would also strengthen the hand of progressive US states and cities that produce low-carbon goods. And it would get attention from targeted countries: US leadership has already shown that it is sensitive to trade measures, as evidenced by its quick reaction to retaliatory tariffs. Critics argue that BCAs are difficult to implement. They point to legal risks and the complexity of measuring carbon footprints.

LEGAL ISSUES

Some policymakers worry that BCAs violate a fundamental principle of the World Trade Organization (WTO): non-discrimination. This principle limits the ability of trade partners to adopt measures that distinguish between equivalent domestic and imported products. For example, it forbids regulators from applying a fee to imported cement unless the same fee applies to locally produced cement. That helps to avert protectionism. It also prevents countries from arbitrarily favouring one trade partner over another.

Because BCAs distinguish between products on the basis of their carbon footprint, they risk being considered discriminatory. For example, producers that make steel in open-hearth or blast-oxygen furnaces might incur a charge; those that produce the alloy in more-efficient electric-arc furnaces might not. Similarly, imposing BCAs on trade partners according to how much carbon they

produce could be seen as favouring some countries over others.

However, because BCAs aim to mitigate climate change, they would fall under a set of exceptions set out in WTO law. These justify measures "necessary to protect human, animal or plant life or health" or "relating to the conservation of exhaustible natural resources". To meet those criteria, BCA

documentation must spell out the environmental goal, and the fees must be imposed through a transparent and fair pro-

"Some of the United States' trade partners could forge a coalition."

cess. Using the revenue to fund mitigation and adaptation efforts, including in developing countries, would strengthen their environmental justification (M. Grubb, *Clim. Policy* **11**, 1050–1057; 2011).

NEXT STEPS

BCAs should be introduced concurrently by a group of countries with the same goals. Some of the United States' trade partners could forge a coalition, including, for example, the EU, Canada and Mexico. Likewise, China has been strongly hit by US tariffs, and has expressed its intention to seek new alliances on climate and trade policy. Together, these nations have considerable economic clout — enough to secure US attention.

As a first step, these countries could base their response to US tariffs on the carbon intensity of goods. That alone would send a message about the importance of climate change.

As these countries advance increasingly ambitious climate policies, they should transition to a BCA that stands on independent footing from the current tariff conflict. The design of the programme must balance legal durability, ease of implementation and environmental performance. Below are our recommendations, based on a research project concluded last year (see go.nature.com/2kdhejm).

Determine scope and coverage. BCAs should target only countries that are not a party to the Paris Agreement. If the United States follows through with its announced withdrawal, its trade partners could impose a BCA. Periodic reviews of the adjustments should be tied to the global stocktaking process under the Paris Agreement. BCAs should be applied to a limited list of imported goods from sectors that emit a lot of carbon, such as iron, steel, aluminium, oil and gas refining, cement and lime, basic inorganic chemicals and pulp and paper. That would reduce the administrative burden, yet still realize many of the environmental benefits. In the United States, these sectors account for roughly half of all manufacturing emissions.

Calculate the footprint and set the adjustment. The carbon footprint should include direct emissions from production and indirect emissions from energy and heat inputs, and could be based on average benchmarks

COMMENT

for each sector. Producers of low-carbon goods should be allowed to document their actual emissions with audited data. The BCA should reflect the difference

in carbon constraints between the imposing and targeted economies. For example, if aluminium in the imposing country faces an average compliance cost of \$30 per tonne of

"Turn this incipient trade war into an opportunity to ratchet up climate ambition."

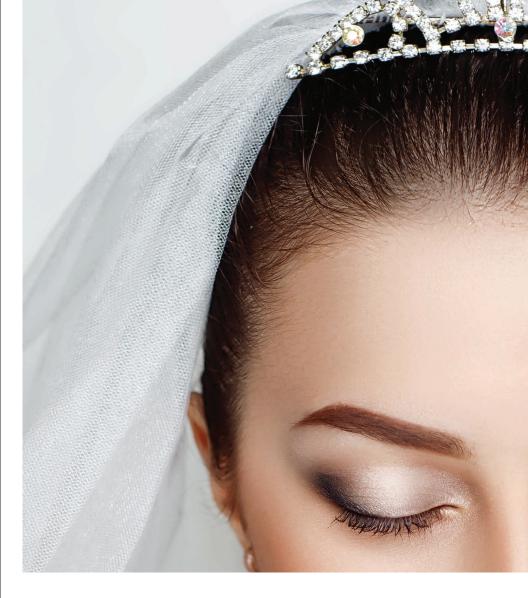
carbon emissions, and imported aluminium is subject only to an average of \$10 per tonne in its country of origin, then the \$20 difference would be imposed as a BCA.

Ensure the process is fair. Equity, transparency and predictability are essential to legal durability. Countries should notify trade partners in advance and discuss the details with them. An independent arbiter could audit the plans before they are adopted and determine whether they are reasonable. The plan should include procedures for appealing.

The next few years will be a crucial time for both trade and climate policy. Trump plans to renegotiate the North American Free Trade Agreement (NAFTA), having already ended US participation in the Trans-Pacific Partnership (TPP). The United Kingdom and EU must rewrite their joint policies on trade and climate change. Nations will review their Paris pledges with a view to strengthening climate ambition. All these processes are parts of a larger puzzle.

As the pieces of the jigsaw fall into place, momentum must be sustained on climate change. Rather than prolonging the current spiral of tariff tit-fortat, countries should rally and turn this incipient trade war into an opportunity to ratchet up climate ambition.

Michael A. Mehling is deputy director of the Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, Cambridge, USA; and professor at the University of Strathclyde School of Law, Glasgow, UK. Harro van Asselt is senior research fellow at the Stockholm Environment Institute; and professor of climate law and policy, University of Eastern Finland Law School, Joensuu, Finland. Kasturi Das is professor of economics and international business, Institute of Management Technology, Ghaziabad, India. Susanne **Droege** is senior fellow, Global Issues Research Division, German Institute for International and Security Affairs, Berlin, Germany. e-mail: mmehling@mit.edu



Design AI so that it's fair

Identify sources of inequity, de-bias training data and develop algorithms that are robust to skews in data, urge **James Zou** and **Londa Schiebinger**.

hen Google Translate converts news articles written in Spanish into English, phrases referring to women often become 'he said' or 'he wrote'. Software designed to warn people using Nikon cameras when the person they are photographing seems to be blinking tends to interpret Asians as always blinking. Word embedding, a popular algorithm used to process and analyse large amounts of natural-language data, characterizes European American names as pleasant and African American ones as unpleasant.

These are just a few of the many examples uncovered so far of artificial

intelligence (AI) applications systematically discriminating against specific populations.

Biased decision-making is hardly unique to AI, but as many researchers have noted¹, the growing scope of AI makes it particularly important to address. Indeed, the ubiquitous nature of the problem means that we need systematic solutions. Here we map out several possible strategies.

SKEWED DATA

In both academia and industry, computer scientists tend to receive kudos (from publications to media coverage) for training ever more sophisticated algorithms. Relatively



little attention is paid to how data are collected, processed and organized.

A major driver of bias in AI is the training data. Most machine-learning tasks are trained on large, annotated data sets. Deep neural networks for image classification, for instance, are often trained on ImageNet, a set of more than 14 million labelled images. In natural-language processing, standard algorithms are trained on corpora consisting of billions of words. Researchers typically construct such data sets by scraping websites, such as Google Images and Google News, using specific query terms, or by aggregating easy-to-access information from sources such as Wikipedia. These data sets are then annotated, often by graduate students or through crowdsourcing platforms such as Amazon Mechanical Turk.

Such methods can unintentionally produce data that encode gender, ethnic and cultural biases.

Frequently, some groups are over-represented and others are under-represented. More than 45% of ImageNet data, which fuels research in computer vision, comes from the United States², home to only 4% of the world's population. By contrast, China and India together contribute just 3% of ImageNet data, even though these countries

represent 36% of the world's population. This lack of geodiversity partly explains why computer vision algorithms label a photograph of a traditional US bride dressed in white as 'bride', 'dress', 'woman', 'wedding', but a photograph of a North Indian bride as 'performance art' and 'costume'².

In medicine, machine-learning predictors can be particularly vulnerable to biased training sets, because medical data are especially costly to produce and label. Last year, researchers used deep learning to identify skin cancer from photographs. They trained their model on a data set of 129,450 images, 60% of which were scraped from Google Images³. But fewer than 5% of these images are of dark-skinned individuals, and the algorithm wasn't tested on dark-skinned people. Thus the performance of the classifier could vary substantially across different populations.

Another source of bias can be traced to the algorithms themselves.

A typical machine-learning program will try to maximize overall prediction accuracy for the training data. If a specific group of individuals appears more frequently than others in the training data, the program will optimize for those individuals because this boosts overall accuracy. Computer scientists evaluate algorithms on 'test' data sets, but usually these are random sub-samples of the original training set and so are likely to contain the same biases.

Flawed algorithms can amplify biases through feedback loops. Consider the case of statistically trained systems such as Google Translate defaulting to the masculine pronoun. This patterning is driven by the ratio of masculine pronouns to feminine pronouns in English corpora being 2:1. Worse, each time a translation program defaults to 'he said', it increases the relative frequency of the masculine pronoun on the web — potentially reversing hard-won advances towards equity⁴. The ratio of masculine to feminine pronouns has fallen from 4:1 in the 1960s, thanks to large-scale social transformations.

TIPPING THE BALANCE

Biases in the data often reflect deep and hidden imbalances in institutional infrastructures and social power relations. Wikipedia, for example, seems like a rich and diverse data source. But fewer than 18% of the site's biographical entries are on women. Articles about women link to articles about men more often than vice versa, which makes men more visible to search engines. They also include more mentions of romantic partners and family⁵.

Thus, technical care and social awareness must be brought to the building of data sets for training. Specifically, steps should be taken to ensure that such data sets are diverse and do not under represent particular groups. This means going beyond convenient classifications — 'woman/man', 'black/white', and so on — which fail to capture the complexities of gender and ethnic identities.

Some researchers are already starting to work on this (see *Nature* **558**, 357–360; 2018). For instance, computer scientists recently revealed that commercial facial recognition systems misclassify gender much more often when presented with darker-skinned women compared with lighter-skinned men, with an error rate of 35% versus 0.8% (ref. 6). To address this, the researchers curated a new image data set composed of 1,270 individuals, balanced in gender and ethnicity. Retraining and fine-tuning existing faceclassification algorithms using these data should improve their accuracy.

To help identify sources of bias, we recommend that annotators systematically label the content of training data sets with standardized metadata. Several research groups are already designing 'datasheets' that contain metadata and 'nutrition labels' for machine-learning data sets (http://datanutrition.media.mit.edu/).

Every training data set should be accompanied by information on how the data were collected and annotated. If data contain information about people,

then summary statistics on the geography, gender, ethnicity and other demographic information should be provided (see 'Image power'). If the data labelling is done through crowdsourcing, then basic information about the crowd participants should be included, alongside the exact request or instruction that they were given.

As much as possible, data curators should provide the precise definition of descriptors tied to the data. For instance, in the case of criminal-justice data, appreciating the type of 'crime' that a model has been trained on will clarify how that model should be applied and interpreted.

BUILT-IN FIXES

Many journals already require authors to provide similar types of information on experimental data as a prerequisite for publication. For instance, Nature asks authors to upload all microarray data to the open-access repository Gene Expression Omnibus — which in turn requires authors to submit metadata on the experimental protocol. We encourage the organizers of machine-learning conferences, such as the International Conference on Machine Learning, to request standardized metadata as an essential component of the submission and peer-review process. The hosts of data repositories, such as OpenML, and AI competition platforms, such as Kaggle, should do the same.

Lastly, computer scientists should strive to develop algorithms that are more robust to human biases in the data.

Various approaches are being pursued. One involves incorporating constraints and essentially nudging the machinelearning model to ensure that it achieves equitable performance across different subpopulations and between similar individuals⁸. A related approach involves changing the learning algorithm to reduce its dependence on sensitive attributes, such as ethnicity, gender, income — and any information that is correlated with those characteristics9.

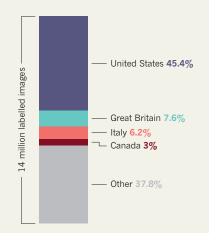
Such nascent de-biasing approaches are promising, but they need to be refined and evaluated in the real world.

An open challenge with these types of solutions, however, is that ethnicity, gender and other relevant information need to be accurately recorded. Unless the appropriate categories are captured, it's difficult to know what constraints to impose on the model, or what corrections to make. The approaches also require algorithm designers to decide a priori what types of biases they want to avoid.

A complementary approach is to use machine learning itself to identify and quantify bias in algorithms and data. We call this conducting an AI audit, in which

IMAGE POWER

Deep neural networks for image classification are often trained on ImageNet. The data set comprises more than 14 million labelled images, but most come from just a few nations.



the auditor is an algorithm that systematically probes the original machine-learning model to identify biases in both the model and the training data.

An example of this is our recent work using a popular machine-learning method called word embedding to quantify historical stereotypes in the United States. Word embedding maps each English word to a point in space (a geometric vector) such that the distance between vectors captures semantic similari-

ties between corresponding words. It captures analogy relations, such as 'man' is to 'king' as 'woman' is to 'queen'. We developed an algorithm — the AI auditor to query the word embedding for

"Biases in the data often reflect deep and hidden imbalances in institutional infrastructures and social power relations.

other gender analogies. This has revealed that 'man' is to 'doctor' as 'woman' is to 'nurse', and that 'man' is to 'computer programmer' as 'woman' is to 'homemaker'1.

Once the auditor reveals stereotypes in the word embedding and in the original text data, it is possible to reduce bias by modifying the locations of the word vectors. Moreover, by assessing how stereotypes have evolved, algorithms that are trained on historical texts can be debiased. Embeddings for each decade of US text data from Google Books from 1910 to 1990, reveal, for instance, shocking and shifting attitudes towards Asian Americans. This group goes from being described as 'monstrous' and 'barbaric' in 1910 to 'inhibited' and 'sensitive' in 1990 — with abrupt transitions after the Second World War and the immigration waves of the 1980s10

GETTING IT RIGHT

As computer scientists, ethicists, social scientists and others strive to improve the fairness of data and of AI, all of us need to think about appropriate notions of fairness. Should the data be representative of the world as it is, or of a world that many would aspire to? Likewise, should an AI tool used to assess potential candidates for a job evaluate talent, or the likelihood that the person will assimilate well into the work environment? Who should decide which notions of fairness to prioritize?

To address these questions and evaluate the broader impact of training data and algorithms, machine-learning researchers must engage with social scientists, and experts in the humanities, gender, medicine, the environment and law. Various efforts are under way to try to foster such collaboration, including the 'Human-Centered AI' initiative that we are involved in at Stanford University in California. And this engagement must begin at the undergraduate level. Students should examine the social context of AI at the same time as they learn about how algorithms work.

Devices, programs and processes shape our attitudes, behaviours and culture. AI is transforming economies and societies, changing the way we communicate and work and reshaping governance and politics. Our societies have long endured inequalities. AI must not unintentionally sustain or even worsen them. ■

James Zou is assistant professor of biomedical data science and (by courtesy) of computer science and of electrical engineering, Stanford University, California, USA. Londa Schiebinger is the John L. Hinds Professor of History of Science and director of Gendered Innovations in Science, Health & Medicine, Engineering, and Environment, Stanford University, California, USA. e-mails: jamesz@stanford.edu; schieb@stanford.edu

- 1. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A. Adv. Neural Inf. Proc. Syst. 2016, 4349-4357 (2016).
- 2. Shankar, S. et al. Preprint at https://arxiv.org/ abs/1711.08536 (2017)
- Esteva, A. et al. Nature 542, 115-118 (2017).
- Schiebinger, L. et al. (eds) Gendered Innovations in Science, Health & Medicine, Engineering and Environment, Engineering, Machine Translation (2011-2015).
- Wagner, C., Garcia, D., Jadidi, M. & Strohmaier, M. Proc. 9th Int. AAAI Conf. Web Soc. Media 454-463
- 6. Buolamwini, J. & Gebru, T. Proc. Mach. Learn. Res. 81. 1-15 (2018).
- Gebru, T. et al. Preprint at https://arxiv.org/ abs/1803.09010 (2018).
- 8. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. *Proc. 3rd Innov. Theor. Comp. Sci. Conf.* 2012, 214-226 (2012).
- Zemel, R., Wu, Y., Swersky, K., Pitassi, T. & Dwork, C. Proc. 30th Int. Conf. Int. Conf. Mach. Learn. 28, III-325-III-333 (2013).
- 10. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Proc. Natl Acad. Sci. USA 115, E3635–E3644 (2018).



SUMMEI BOOKS

Discover deep and illuminating summer reads picked by our regular reviewers, leagues away from lab and lecture hall.



When Einstein Walked with Gödel

Јім Ногт Farrar, Straus and Giroux (2018)

Here, philosopher Jim Holt gathers two decades' worth of reviews and essays, written for The New York Review of Books, The New Yorker and other publications. These are bold, thought-provoking pieces that tackle everything from the physical principle of least action to the mathematics of the infinitely great and infinitesimally small. As a physicist, I was aware that the zeros of the zeta function describe the distribution of the prime numbers. The related Riemann hypothesis is one of the great unsolved maths problems, and Holt helped me to understand why it's important. If true, there's a hidden harmony to the primes; if false, many supposed proofs of contemporary mathematics would fall with it.

There is some repetition of Holt's obsessions. These include the dichotomy between philosophies of mathematics, and the logician Kurt Gödel's claim (during his citizenship hearing) that the US Constitution was logically inconsistent. Holt has revised some pieces. A discussion of Henry Frankfurt's On Bullshit, dating from the first flowering of "truthiness" in the administration of president George W. Bush, is updated with a depressing reference to Donald Trump.

These are stories of real humans and their mathematical, physical and philosophical theories — some of the most complex ever devised.

Andrew Jaffe is a cosmologist and head of astrophysics at Imperial College London.



Health Outcomes in a Foreign Land

Bernard Kwabi-Addo Springer (2017)

In any society, many interconnected factors contribute to disparities in chronic noninfectious diseases such as diabetes. In a country as complex as the United States, identifying and understanding those interactions is daunting. Cancer researcher Bernard Kwabi-Addo has made this effort, uncovering a vast territory of associated variables with important ramifications for health.

This is a multilayered, interdisciplinary survey of genetic and non-genetic influences on health inequities, mainly among Americans of African descent. It offers a wealth of comparative statistics on US citizens, although based on official racial and ethnic categories, which are stereotypical and imprecise. Intertwined is a literature-based social commentary on environmental contributions to discrepancies in health outcomes, mainly in cardiovascular disease, obesity, diabetes, end-stage renal disease and hypertension and various cancers.

This is not a deep anthropological study with original solutions or historically contextualized analyses of the health of African Americans. Rather, it is a highly readable epidemiological synopsis from an under-represented viewpoint — that of a recent West African immigrant confronting a country founded by immigrants. As such, it offers a new, potential source of insights.

Fatimah Jackson is professor of biology and director of the W. Montague Cobb Research Laboratory at Howard University in Washington DC.



Darwin and the Making of Sexual Selection

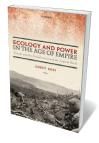
EVELLEEN RICHARDS Univ. Chicago Press (2017)

In 1871, Charles Darwin published The Descent of Man, and Selection in Relation to Sex. Sexual selection solved evolutionary puzzles that natural selection could not. Generation after generation, peahens chose peacocks with the most impressive plumage, producing the peacock's uncamouflageable tail. Darwin believed that, in a similar way, sexual selection (driven by mates selecting according to culturally specific beauty standards) unlocked the mystery of human diversity's origins.

It troubled Darwin, a privileged white Victorian man, to impute agency to women and aesthetic discrimination to non-Europeans. His peers rejected the theory. But biologists are revisiting it. Science historian Evelleen Richards's book vividly excavates its origins.

Darwin developed his ideas on sexual selection while immersed in fields as diverse as embryology and pigeon breeding. Deeply personal matters such as choosing his wife, Emma, and daily preoccupations such as women's fashions, also played a part. In Richards's view, Darwin's opposition to slavery did not, as others argue, motivate his work on sexual selection. What did was his human attempt to answer scientific, political, social and personal questions.

Elizabeth Yale is a lecturer in the history department at the University of Iowa in Iowa City. She is the author of Sociable Knowledge.



Ecology and Power in the Age of Empire

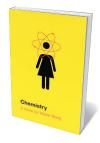
COREY ROSS Oxford Univ. Press (2017)

More and more people now are 'resource omnivores', able to draw the stuff of their lives whether smartphones or kiwi fruit — from across the globe. That fundamental change in the human relationship with nature is often taken for granted. But this chronicle by historian Corey Ross challenges us to acknowledge the imperial roots of modern consumer culture.

In the nineteenth and twentieth centuries, Europeans transformed much of tropical Asia and Africa through vast plantations, mining operations and infrastructure construction to support their expanding mass-production economies. Ross's brilliant book shows how complex and far-reaching that transformation was. The exploitation of cotton, cocoa, rubber, tin and copper led many colonizers to worry about environmental degradation. It also made 'development' a nearly universal goal for post-colonial societies.

Ross concludes that the inequities of empire add to the difficulty of solving contemporary environmental problems. So do the habits of thought that empire spread, especially the idea that the world is an inexhaustible storehouse of resources to be grabbed and exploited in the quest for ever-greater wealth and power.

Adam Rome teaches history at the University at Buffalo, New York. He is writing the environmental-history volume for Oxford's Very Short Introduction series.



Chemistry

Weike Wang Knop (2017)

Awareness of the mental-health issues affecting many PhD students is now widespread. So a novel exploring depression during graduate studies is timely especially one so finely wrought as Weike Wang's Chemistry.

The unnamed Chinese American protagonist of this debut novel is entrapped by parental expectation, a commitment-obsessed boyfriend and a stubbornly unsuccessful doctoral project in synthetic organic chemistry. She doesn't seem to realize how badly she wants out of all three until the day she finds herself hurling glassware onto the lab floor. The novel investigates how she got to that point psychologically, and details her attempts to extract herself to find a better life beyond.

Despite the bleak topic, Wang's anonymous protagonist is luminously funny. There wasn't a vignette where I didn't find myself laughing out loud at the gallows humour — or ruefully shaking my head at thorny issues that all scientists will recognize. Those range from the mundane (non-stop experimental failure) to the extreme (lab heads who drive their students so ruthlessly that they resort to fraud, or even suicide). When a student tragically kills himself in the novel, he is described as "considerate" because he left his flatmates a note on his body: "Danger: potassium cyanide. Do not resuscitate."

Jennifer Rohn is a cell biologist at University College London, editor of the webzine LabLit. com, and writer. Her latest novel is Cat Zero.



Tell Me the Planets

BEN PLATTS-MILLS Fig Tree (2018)

Confabulation, ataxia, dysarthria, dysphagia, hemiparesis: the mesmerizing names of neurological conditions mask cruelties. But in Ben Platts-Mills's extraordinary book, the characters of people damaged by violence, stroke or accident of birth outshine the medical details.

The narrative begins in a London charity called Headway. Through its doors come individuals who have maintained their identity, however uncertainly. Danny is half paralysed after a gangland assault; Liah, born infected with HIV, has been left almost helpless by a brain-damaging condition diagnosed too late; ambitious computer programmer Matthew found his life redirected when a colloidalcyst operation led to memory impairment.

The three are an ex-convict Londoner, an Eritrean refugee and a Nigerian economic migrant. But what really defines these individuals, even as memory fails and words elude them, is their stubborn vitality, their awareness of their own condition as they face bureaucracy, social and medical. They know, and they want us to know.

. 'How do you tell a story about so much loss, about disability, about catastrophic misfortune ... ?" Platts-Mills asks. He tells it wonderfully.

Tim Radford was until 2005 science editor of The Guardian. His book The Consolations of Physics (or, The Solace of Quantum) will be published in August.



Pain, Pleasure. and the Greater Good

CATHY GERE Univ. Chicago Press (2017)

In this thoroughly gripping science history of utilitarianism, Cathy Gere charts the trajectory of the ethical theory, which hinges on the 'greatest good for the greatest number'. For 200 years, utilitarianism pervaded much research in medicine and psychology: pain inflicted on individuals was seen as providing broader gain and a platform for social policy. Reducing aid to the poor, for instance, was understood to 'save' society from the indolence this supposedly encouraged.

Gere's engrossing narrative takes us up to the 1973 hearings on the notorious Tuskegee Syphilis Study. For four decades, the US Public Health Service had observed the progression of the disease in hundreds of impoverished African American men, who were neither told they carried it nor given treatment. Medical claims of greater good were brought crashing down. Yet the study's ethos resurfaces in behavioural economics, through nudges that, without consent, shape the many in the mould of the few — supposedly 'saving' us from some inherent irrationality. Gere rightly emphasizes that we should be wary of 'noble' ends justifying any means.

Alex Haslam is professor of psychology and Australian Laureate Fellow at the University of Queensland in Brisbane. His most recent book, which he co-authored, is The New Psychology of Health.



Adaptive Markets

Andrew Lo Princeton Univ. Press (2017)

The idea that financial markets are in any way rational or efficient seems, to many, absurd - not least as we mark the tenth anniversary of the 2008 crisis. Yet in economics, Eugene Fama's 'efficient-market hypothesis' has a stubborn grip. This holds that prices of financial assets incorporate all available information, so that those assets will always trade at their objectively justified value. You can see the point: if there were unexploited information, somebody would trade on it and thus incorporate it into the market price.

In this study, economist Andrew Lo proposes an alternative: the adaptive-market hypothesis. In stable times, he argues, the efficient-market insight holds; but when, say, a shock bank collapse occurs, other kinds of investor behaviour, such as panic selling, can kick in. Lo draws extensively on neuroscience, psychology and evolutionary biology to develop his ambitious theory.

As he acknowledges, there is a lot still to understand. But it's a fascinating read, cogently situating financial behaviour within what we know about human behaviour and evolutionary history. What's more, Lo concludes, his theory can inform the regulation of financial markets so that they serve society by funnelling investment to tackle big problems such as climate change — rather than the socially destructive shortterm trading that led to the 2008 meltdown.

Diane Coyle is Bennett Professor of Public Policy at the University of Cambridge, UK.



Correspondence

Trump cannot crush Iran's scientists

US President Donald Trump's unilateral withdrawal from the Iran Nuclear Agreement (Joint Comprehensive Plan of Action; ICPOA) in May attracted international condemnation. As vice-dean for research in the Faculty of Medicine at Tehran University of Medical Sciences, I stand behind Iran's scientists, who have resolved to work even harder to maintain the country's scientific progress (see also Nature 557, 287-288; 2018).

After the imposed war in 1980-88 and decades of Western sanctions, Iran has made remarkable advances in research. ranking 17th in the world in 2012. The JCPOA did not have much impact on scientific productivity, in part because many US sanctions remained in place. These continued to affect the purchase of books, journals, lab equipment and materials; the payment of publication charges; membership of scientific bodies; and travel to conferences and meetings. Furthermore, the US treasury department clamped down on publication in US journals of papers from Iranian government scientists (see S. Akhondzadeh Avicenna J. Med. Biotechnol. 5, 203; 2013).

In the face of Trump's withdrawal from the JCPOA, I hope that the international scientific community will support Iran's efforts to contribute further to international science.

Shahin Akhondzadeh Tehran University of Medical Sciences,

s.akhond@tums.ac.ir

Exit interviews and lab-member awards

As leader of a large research group, I would like to share an effective strategy for collecting negative feedback and constructive suggestions from lab members on leadership issues (see Nature 557, 294-296; 2018).

Following the practice of

many commercial companies, I organize an exit interview with every postdoc, graduate and undergraduate student when they leave the lab. I find that people are generally more open about problems when they are leaving, because they no longer have to worry about reactions from their seniors or colleagues. Identifying likes and dislikes from a variety of viewpoints helps me to reinforce good practices and modify unwelcome ones.

Another industrial ploy I use is to run semi-annual votes for the best lab member. along the lines of company awards for 'employee of the month'. Lab members vote on three performance criteria: helpfulness, work ethic and productivity. The person who obtains the highest collective score from their peers is treated to a free lunch.

Although the winners value their peers' respect over a free lunch, the award helps the lab establish a culture of helping one another, working hard and with integrity, and honing scientific findings for publication. Z. Hugh Fan University of Florida, Gainesville, Florida,

e-mail: hfan@ufl.edu

Evaluation woes: we saw it coming

The cry of anguish from John Tregoning asking how his research should be judged, if not by the journal impact factor (Nature 558, 345; 2018), reflects a profound malaise in the university system. So what did we do before journal impact factors were invented, when career advancement flourished anyway?

The transition from traditional rigorous intellectual assessment of research to bibliometric indices and box-ticking coincided with the transition to the corporate university model and the rise of the university bureaucrat. These administrators showed less interest in assessing the intellectual merit of research

than in deploying competitive metrics for the marketplace.

Governments are much to blame because of their decreasing budgets for tertiary education. However, the professoriate (to which I belong) should have seen the danger these shifts posed sooner and, when it did, it should have fought harder for the intellectual heart of the system.

Some evidence-based metrics are useful. In my view, however, a return to the methods of peerdriven intellectual assessment that worked well for centuries should remain part of the answer to evaluation woes even though that could mean retrieving the system from the grasp of university bureaucrats and the burgeoning bibliometric

Andrew Beattie Macquarie University, Sydney, Australia. andrew.beattie@mq.edu.au

Evaluation woes: start right

In our view, we need to move from a single system for assessing research performance (see J. Tregoning Nature 558, 345; 2018) to a prospective model implemented at the start of a research initiative. This would engage stakeholders in defining metrics for the project's mission and agenda.

An example is the European Commission's MULTI-ACT project, which is a collective research-impact framework of multivariate models for health research and innovation (see go.nature.com/2mdkqgt). This integrates conventional metrics related to excellence with new measures relating to economic and financial efficiency and to social efficacy.

Although not the "quick fix" Tregoning mentions, such multidimensional measures should help early-career researchers to tie their work more effectively to a meaningful research agenda.

Paola Zaratin Italian Multiple

Sclerosis Foundation, Genoa, Italy. Marco Salvetti NESMOS, Sapienza University of Rome, IRCCS Neuromed, Italy. paola.zaratin@aism.it

Evaluation woes: metrics beat bias

We disagree with the contention that publication metrics should be condemned as the bane of research-evaluation practices (see J. Tregoning Nature 558, 345; 2018). In countries with a long-rooted tradition of nepotism and patronage, such metrics provide objective and consistent evaluation particularly advantageous for early-career researchers. They can also help overstretched funding agencies and review panels to arrive at fast, fair and transparent decisions.

The conventional combination of qualitative review and quantitative metrics can be expensive and time-consuming, not least because it is hard to find genuinely impartial reviewers and to achieve consensus.

We acknowledge that misuse of metrics such as journal impact factors and citation counts can discredit creative research, encourage citation gaming and provoke research misconduct. But the striking increase in the popularity of metrics as an evaluation tool worldwide indicates that they offer benefits, too.

Li Tang Fudan University, Shanghai, China. Guangyuan Hu Shanghai

University of Finance and Economics, China. litang@fudan.edu.cn

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature. com after consulting the author guidelines and section policies at http://go.nature.com/ cmchno.

NEWS & VIEWS

IMAGING TECHNIQUES

A record-breaking microscope

An electron microscope has been developed that produces images at higher resolution than conventional approaches can achieve, and is suitable for studying fragile materials that can be damaged by electron beams. SEE ARTICLE P.343

JOHN RODENBURG

n page 343, Jiang et al. 1 report the highest-magnification image ever obtained using a transmission electron microscope. The image reveals the atoms in a two-dimensional self-supporting sheet of a semiconductor, and has a resolution of 0.39 ångströms; for comparison, most atoms are about 2–4 Å in diameter. The technique might eventually allow 2D materials to be examined with unprecedented precision, providing insight into this burgeoning class of useful compounds. It might also lead to the development of a method that can image individual atoms in 3D objects.

To generate their image, the authors used a method called ptychography (the 'p' is silent) in which radiation — in this case, electrons — is passed through a specimen to produce many 2D diffraction patterns. The basic principle of the technique was proposed almost 50 years ago by the physicist Walter Hoppe, who reasoned that there should be enough information in the diffraction data to work backwards to produce an image of the diffracting object². However, it was many years before computer algorithms were developed that could do this reverse calculation easily and reliably^{3,4}. The pictures produced by ptychographic methods are generated using a computer from a vast amount of indirect scattering data. An important advantage of this approach over conventional microscopy is that it can surpass the resolution limit imposed by lens imaging. In fact, it can work without any lenses at all⁵.

Over the past ten years, ptychography has been widely used for microscopy in the X-ray⁴, extreme ultraviolet⁶ and visible-light^{7,8} regions of the electromagnetic spectrum. It has also been used with some success with electrons, but Jiang and colleagues are the first to show that it can surpass the resolution obtained by the best electron lenses. For the technique to work, every electron must be counted almost perfectly, but the scattering patterns contain both extremely bright and extremely dark regions (they are said to have a high dynamic range), which makes it difficult to count every electron without any errors. To complicate matters further, the experiment must be done as quickly as possible to fulfil other experimental constraints, which means that about 1,000 diffraction patterns must be recorded every second.

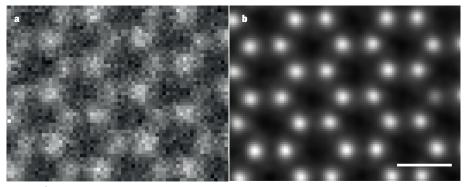


Figure 1 | **Improving the resolution of electron microscopy. a**, This image of a sheet of molybdenum disulfide was obtained using annular dark-field electron microscopy — the conventional method for obtaining extremely high-resolution images of samples. **b**, Jiang *et al.*¹ report an electron microscope that works by analysing the diffraction patterns of electrons that have been transmitted through a sample (a technique known as ptychography). This method provides the best resolution yet reported for an electron microscope. Here, the atoms in the sheet of molybendum disulfide are much clearer. Scale bar, 3 ångströms.

Speed, accuracy and dynamic range are conflicting performance parameters in any electron detector — achieving them all is difficult. All previous electron detectors have compromised on one or more of these properties. The authors' main achievement is to implement a detector that can handle such demanding specifications.

Remarkably, Jiang et al. gave themselves a huge handicap with regard to beating the resolution record. For any given microscope lens, the best resolution is achieved by using the shortest possible wavelength of the radiation or electron beams concerned. However, the authors used relatively low-energy electrons, which have twice the wavelength of those used in the highest-resolution lens-based microscopes^{9,10}. Using low-energy electrons for microscopy is good because it greatly reduces the damage inflicted on the specimen by the electrons. But in this case, it also meant that the resolution of the lens used by Jiang and colleagues was reduced by a factor of two. To beat the resolution record, the authors had to process a particular subset of the ptychographic diffraction data (the high-angle data), thereby obtaining an image with a resolution 2.5 times better than would otherwise have been possible.

Achieving high resolution is not the whole story, however. Anyone with poor eyesight knows that they need as much light as possible if they want to read small print. This is because there is an intimate relationship between

resolution, contrast and the amount of light illuminating the object. If the small print is light grey, not black, then the contrast is low, and even more light is needed to read it. The same principle is true for an electron microscope.

Jiang and colleagues used ptychography to work out how a particular property of the electron waves, known as the phase, changes as the waves pass through an object. This information can be used to produce images that have strong contrast — even for specimens that contain atoms of low atomic number, which are difficult to detect with conventional electronmicroscopy methods that offer very high resolution. The authors therefore needed relatively few electrons to generate their images compared with other state-of-the-art techniques, such as annular dark-field electron microscopy (Fig. 1). So not only did they use low-energy electrons without compromising resolution, but they also used many fewer electrons than other techniques do, further reducing the damage done to the sample.

Perhaps the most striking feature of Jiang and co-workers' image is not the atoms themselves but the enormous gaps between them. The average bond lengths in a material can be measured in a bulk sample by using all sorts of diffraction and spectroscopic methods, but the authors' image provides an extremely precise measurement of the lengths of the bonds between individual pairs of atoms, which are sensitive to the atoms' local bonding

environment. But are ultrahigh-resolution images of gaps between atoms useful for anything else?

I think the answer lies in the big success story of X-ray ptychography: tomography⁴, a technique in which lots of 2D images of a transparent object are acquired as it is rotated, so that a 3D image can be built up. Phase information is an ideal imaging signal for this technique. But when images are taken through a solid object, the resolution needs to be as high as possible to distinguish features lying on the top surface from those at the bottom, many of which will seem (when seen in projection) to be laterally close to one another.

Jiang *et al.* tested the resolution of their electron microscope by putting two layers of atoms on top of one another and measuring the minimum apparent lateral distance between atoms in different layers, some of which were almost overlapping. In my view, this test demonstrates that their instrument could potentially be used for tomography. In

theory, such imaging of multiple layers is not limited to crystalline 2D materials and could be used for any complicated, non-crystalline structure. Unfortunately, for thicker objects, the electron waves would scatter so strongly that they would spread out and re-interfere with each other in complicated ways, which would make it even harder — although in theory not impossible — to work out the structure.

Perhaps the take-home message of this work is not so much the record resolution, or its applications to 2D materials, but the fact that it will provide a way of precisely imaging the 3D bonding of every individual atom in a solid volume of matter, while using a minimal flux of damaging electrons. Indeed, the authors allude to this enticing possibility in their conclusions, suggesting that the next step is to use their remarkable detector for tomography. The aim would then be to solve the exact 3D atomic structures of solids that have no longrange order, such as nanocrystalline materials, glasses and amorphous metals, for which we

must currently infer structures from averaged bulk measurements. ■

John Rodenburg is in the Department of Electronics and Electrical Engineering, University of Sheffield, Sheffield S10 2TN, UK. e-mail: j.m.rodenburg@shef.ac.uk

- 1. Jiang, Y. et al. Nature 559, 343–349 (2018).
- 2. Hoppe, W. Acta Crystallogr. A **25**, 495–501 (1969).
- 3. Rodenburg, J. M. in *Advances in Imaging and Electron Physics* Vol. 150 (ed. Hawkes, P. W.) 87–184 (Elsevier, 2008).
- 4. Pfeiffer, F. Nature Photon. 12, 9-17 (2018).
- Rodenburg, J. M., Hurst, A. C. & Cullis, A. G. Ultramicroscopy 107, 227–231 (2007).
- Gardner, D. F et al. Nature Photon. 11, 259–263 (2017).
- Marrison, J., Räty, L., Marriot, P. & O'Toole, P. Sci. Rep. 3, 2369 (2013).
- Zheng, G., Horstmeyer, R. & Yang, C. Nature Photon. 7, 739–745 (2013).
- Erni, R., Rossell, M. D., Kisielowski, C. & Dahmen, U. Phys. Rev. Lett. 102, 096101 (2009).
- 10.Morishita, S. *et al. Microscopy* **67**, 46–50 (2018).

The author declares competing financial interests See go.nature.com/2nablvu for details.

CARDIOVASCULAR BIOLOGY

Cells stop dividing to become arteries

An analysis of gene-expression patterns in single cells provides detailed insights into the developmental processes that lead to maturation of the coronary arteries. SEE ARTICLE P.356

ARNDT F. SIEKMANN

The human heart pumps between about 5 and 20 litres of blood through the body every minute1. To receive enough oxygen to fulfil this tremendous task, heartmuscle cells need their own blood supply. This is provided by specialized blood vessels, including coronary arteries. Defects in these arteries can lead to coronary heart disease and even heart attack^{2,3}. Understanding how coronary arteries form during embryonic development is therefore of great interest, because such knowledge might help in developing strategies to prevent or treat coronary heart disease. On page 356, Su et al.4 provide a detailed picture of the sequence of events that leads to coronary artery development.

The cells that generate coronary arteries originate from various regions of the embryo, including a sac-like structure called the sinus venosus that adjoins the embryonic heart^{5,6}. From these sites, the cells invade the heart's muscle-cell layer. Here, they form an immature blood-vessel network called a plexus that is subsequently remodelled into functional arteries and veins.

Su and colleagues set out to investigate

how cells from the sinus venosus develop into coronary arteries, using single-cell RNA sequencing (scRNA-seq) — a technique that enables precise identification of the genes being expressed in each cell of a tissue⁷. Geneexpression patterns change during tissue differentiation, for example as sinus venosus cells mature into coronary arteries. Comparison of the gene-expression patterns for individual cells of a given type can therefore reveal the cells' relationships to one another.

The authors extracted single endothelial cells, which make up the inner lining of blood vessels, from the hearts of mouse embryos at a developmental time point just before coronary artery formation. They reasoned that, at this embryonic stage, they would obtain cells at the various stages leading to coronary artery maturation, including sinus venosus and plexus cells. They then used bioinformatics to investigate the lineage relationships between these cells.

It has been thought that the remodelling of

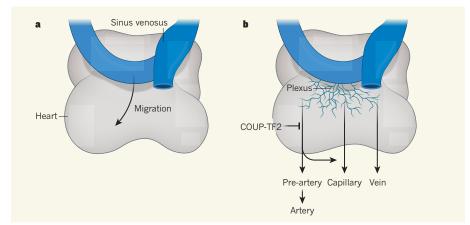


Figure 1 | **Coronary artery development starts early.** a, During the development of mouse embryos, cells from a sac-like structure called the sinus venosus migrate into the muscle-cell layer of the heart. b, There, they give rise to an immature blood-vessel network (a plexus), which will be remodelled to form arteries, veins and capillaries. Su *et al.* have shown that a subpopulation of immature plexus cells, which the authors dub pre-artery cells, have a gene-expression profile that is characteristic of mature arteries. The transcription factor COUP-TF2 prevents plexus cells from adopting this profile. Pre-artery cells predominantly give rise to mature coronary artery cells, although a few become part of capillaries instead. (Figure adapted from Fig. 4h of ref. 4.)

the plexus into arteries and veins starts only after the plexus has connected to the aorta (the main heart artery), and therefore after the onset of blood flow⁵. But, unexpectedly, Su et al. found that several cells from their embryos, in which the plexus had not yet received blood, had a gene-expression profile associated with mature arteries. They called these cells pre-artery cells.

The authors used a genetic strategy to indelibly label the pre-artery cells with a marker protein, such that these cells and the lineages they give rise to could be tracked during embryonic development. This lineage tracing revealed that, although most pre-artery cells did go on to form coronary arteries, some were incorporated into capillaries, which connect coronary arteries with veins. Thus, it seems that, although certain endothelial cells are genetically predisposed to form arteries, they also have a degree of developmental plasticity (Fig. 1).

Next, Su and colleagues performed a detailed analysis of the gene-expression patterns of cells on the developmental spectrum from sinus venosus to pre-artery cells. Changes in gene expression towards more arterial-like profiles occurred only gradually along most of the spectrum. However, there was a sharp change as cells crossed a threshold to adopt a pre-artery state. The researchers showed that the greatest difference in expression in pre-artery cells compared with other cells in their analysis occurred in genes implicated in regulating the cell cycle. Furthermore, in mouse embryos, pre-artery cells proliferated less than did cells in the plexus. Thus, limiting cell divisions might be a prerequisite for coronary artery maturation.

Indeed, the authors found that overexpression of the transcription factor COUP-TF2 in mice inhibited pre-artery formation by upregulating cell-cycle genes. COUP-TF2 was previously thought to limit the growth of arteries by suppressing the Notch signalling pathway8. But Su et al. showed that activation of Notch signalling could not prevent the defects caused by COUP-TF2 overexpression in mouse embryos. By contrast, pharmacological inhibition of the cell cycle increased artery formation in an ex vivo experiment. Thus, COUP-TF2 has functions in artery development that are independent of Notch signalling. Together, Su and colleagues' work provides exciting insights into coronary artery formation. It will be interesting to discover whether the findings apply to artery development in other settings.

It will also be valuable to delineate the signalling pathways that lead certain cells to adopt the gene-expression profile of preartery cells. The Notch signalling pathway, acting independently of COUP-TF2, is a prime candidate. Studies in several developmental settings9-11 have suggested that new blood-vessel sprouts initially emanate from veins and capillaries and only subsequently

form arteries, with each activity inhibiting the other. Inhibiting Notch signalling can lead to excessive blood-vessel sprouting, while at the same time preventing artery formation^{12,13}, and a similar effect is seen during coronary artery development¹⁴. It will therefore be interesting to investigate how Notch signalling affects the gene-expression profiles that lead to the formation of pre-artery cells.

Although it is becoming increasingly clear that artery differentiation is intimately linked to cell-cycle state, the underpinnings of this relationship need further investigation. A report last year showed that cell-cycle inhibition is important for proper arterial gene expression in cells in vitro¹⁵. In addition, the signalling pathways involving Notch and vascular endothelial growth factor, which are indispensable for the establishment of new blood-vessel networks, are both implicated in influencing endothelialcell proliferation¹⁶. However, it is not known how cells interpret signalling inputs from these pathways to balance the demand for proliferation of cells in the blood-vessel network with the need to establish new arteries.

One must also bear in mind that cell-lineage trajectories obtained from scRNA-seq might not reflect true developmental relationships. For instance, cells that have similar gene-expression patterns are not necessarily derived from the same precursor population. New techniques that unite cell-lineage tracing with scRNA-seq¹⁷ will help to bridge this gap, and will surely provide further insights into coronary artery development. ■

Arndt F. Siekmann is at the Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany.

e-mail: arndt.siekmann@mpi-muenster.mpg.de

- 1. Epstein, S. E., Beiser, G. D., Stampfer, M., Robinson, B. F. & Braunwald, E. Circulation 35, 1049-1062 (1967).
- Gisterå, A. & Hansson, G. K. Nature Rev. Nephrol. 13, 368-380 (2017).
- GBD 2016 DALYs and HALE Collaborators. Lancet **390**, 1260–1344 (2017). 4. Su, T. et al. Nature 559, 356–362 (2018).
- Sharma, B., Chang, A. & Red-Horse, K. *Annu. Rev. Physiol.* **79**, 1–19 (2017).
- Tian, X., Pu, W. T. & Zhou, B. Circ. Res. 116, 515-530
- 7. Potter, S. S. Nature Rev. Nephrol. https://doi. org/10.1038/s41581-018-0021-7 (2018).
- You, L.-R. et al. Nature 435, 98-104 (2005). 9. Bussmann, J., Wolfe, S. A. & Siekmann, A. F.
- Development 138, 1717-1726 (2011). 10.Xu, C. et al. Nature Commun. 5, 5758 (2014).
- 11. Laviña, B. et al. Development https://doi.org/ 10.1242/dev.161182 (2018).
- 12. Hasan, S. S. et al. Nature Cell Biol. 19, 928-940
- (2017). 13. Pitulescu, M. E. et al. Nature Cell Biol. **19**, 915–927 (2017).
- 14. Wang, Y. et al. Nature Commun. 8, 578 (2017)
- 15. Fang, J. S. et al. Nature Commun. 8, 2149 (2017) 16. Potente, M. & Mäkinen, T. Nature Rev. Mol. Cell Biol.
- **18**, 477–494 (2017) 17. Kester, L. & van Oudenaarden, A. Cell Stem Cell

This article was published online on 4 July 2018.

Newfound differences between great apes

High-quality genome sequences for some of the great apes have been assembled using state-of-the-art sequencing tools. The assemblies provide an unbiased comparison between humans and their closest evolutionary relatives.

AYLWYN SCALLY

uch of evolutionary biology is motivated by the principle that you cannot understand one species without comparing it with another. When nineteenth-century naturalists compared the anatomies of humans and other apes, it became clear that these species shared many features and had evolved from a common ancestor. More recently, developments in DNA sequencing — which enabled assembly of the human genome¹ in 2001, followed by lower-quality 'draft' genomes for other great apes²⁻⁴ — have transformed our understanding of this evolutionary process. Writing in Science, Kronenberg et al.5 describe new great-ape genome assemblies, generated using a technology that surpasses previous methods.

This work marks a new stage in our ability to study and compare these species.

Genome assembly is often likened to piecing together a jigsaw puzzle — a huge jigsaw for which the box has been lost and we have only a vague idea of what the whole should look like. The analogy holds because sequencing technologies cannot sequence an entire chromosome in one go. Instead, they fragment the genome into many separate pieces, called reads, which have to be matched, overlapped and placed together.

Previous generations of sequencing machines produced reads that were only about a hundred base pairs long, or perhaps a thousand base pairs but at exorbitant cost. Current machines such as Pacific BioScience's singlemolecule real-time (PacBio SMRT) sequencer produce reads tens of thousands of base pairs in length. Even with this improvement, hundreds



Figure 1 | **A Sumatran orangutan and her baby.** Kronenberg *et al.*⁵ assembled high-quality genome sequences for a chimpanzee and an orangutan, and compared these with the human genome to look for evolutionary differences.

of thousands of reads are needed to span a genome of three billion base pairs such as that of humans, Moreover, in practice, a large excess is used (typically more than 30 genomes' worth) to mitigate errors and resolve overlap ambiguities. A further complication arises from the fact that genomes are filled with stretches of DNA in which the same pattern is repeated many times, either in series or scattered throughout the genome. In apes, such repetitive DNA comprises a substantial fraction of the genome.

Because of these difficulties, the first greatape genome projects used the human genome as a scaffold to help assemble genomic regions that are structurally similar to those of humans — that is, in which corresponding stretches of DNA lie in the same order and are present in a similar number of copies. This strategy enabled better assembly in such regions. But in regions where genome structure has evolved very differently in humans and other great apes, the great-ape draft assemblies tended to be more fragmented, and the resulting variation in assembly quality effectively constituted a bias towards the human genome. These assemblies provided many evolutionary insights, but there has nonetheless been a deficit in our understanding of the

genomic elements that make humans unique.

One reason why structural variation is important, particularly on the short evolutionary timescale that separates humans and other great apes, is that it provides a way for genomes to evolve rapidly. When a whole chunk of DNA is removed or duplicated, its molecular function can be inhibited or enhanced in one step, rather than through successive mutations at individual bases. Indeed, much of the

"Structural variation also seems to have had a role in brain evolution."

great-ape genome seems to be modular in nature, and is therefore susceptible to the kind of building-block alteration that structural variation allows. It is also

thought that gene loss is a key mechanism for evolutionary change^{6,7}. This might seem counterintuitive, but genes often act to constrain, rather than promote, a particular function. Disabling them by removing, duplicating or relocating a chunk of DNA might be the simplest way to confer beneficial effects.

Kronenberg *et al.* used PacBio SMRT to assemble high-quality genomes for a chimpanzee and an orangutan, along with two

human genomes for comparison (Fig. 1). The long reads enabled them to do away with the human-genome scaffold used previously, and to increase the typical distance between gaps by about 100-fold compared with previous assemblies. The authors found about 600,000 structural differences between these genomes and that of humans, including more than 17,000 differences specific to humans. Of these, many changes disrupt genes in humans that are not disrupted in other apes. Genes whose activity is suppressed specifically in humans are more likely than other genes to be associated with a human-specific structural variant.

Many genes produce multiple versions, called isoforms, of the protein they encode, each of which can have a different role. Kronenberg and colleagues found evidence that one human-specific structural change — a large deletion in the gene *FADS2* — might have altered the distribution of isoforms the gene produces. These isoforms are involved in the synthesis of fatty acids needed for brain development and immune response⁸, and are difficult to obtain from a purely herbivorous diet. Correspondingly, *FADS2* has been a target for natural selection associated with dietary changes towards or away

from animal fats in recent human evolution⁸. Chimpanzees eat a small amount of meat, so it is not known what (if any) human-specific traits might have resulted from this deletion, but it does suggest that shifting dietary patterns could have been a feature of human evolution over long timescales.

Structural variation also seems to have had a role in brain evolution. Human brains are much larger than those of other apes, and it is plausible that genes involved in brain growth and development were key to the evolution of this trait. The authors analysed the sequences of genes that are active in radial glial cells, which are progenitors for neurons and other cells in the brain's cortex, and compared protein production by these genes in humans and chimpanzees using cortical organoids — 3D models of brain tissue grown in vitro. These analyses revealed that 41% of genes whose activity is suppressed in human radial glial cells are associated with a human-specific structural variant. Again, this

is consistent with structural genomic changes causing disruption or loss of gene function during great-ape evolution.

Intriguing as Kronenberg and colleagues' findings are, there is also a broader significance to their work. Several groups and consortia are applying new sequencing technologies to different organisms. Ultimately, researchers want accurate, high-resolution assemblies for all species, and to compare these genomes on an equal footing. This will improve evolutionary analyses and reveal complex mutation processes that have hitherto been obscured. Large genome assembly currently remains hugely expensive, and even state-of-the-art sequencing tools struggle to resolve repetitive sequences on scales above a few hundred thousand base pairs, making assembly of certain genomes challenging. But tools to read whole genomes with negligible errors on inexpensive hardware are not far away, and are almost available for small bacterial genomes9.

It is clear that we are leaving behind the

initial period of evolutionary genomics, in which analyses involved comparing a genome of interest to a few 'gold standard' genomes, such as human, mouse or zebrafish. Instead, we are moving towards a more complete and equable genomic view of life. ■

Aylwyn Scally is in the Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK.

e-mail: aos21@cam.ac.uk

- 1. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
- Chimpanzee Sequencing and Analysis Consortium. Nature 437, 69–87 (2005).
- 3. Locke, D. P. et al. Nature 469, 529-533 (2011).
- 4. Scally, A. et al. Nature **483**, 169–175 (2012).
- Kronenberg, Z. N. et al. Science 360, eaar6343 (2018).
- 6. Olson, M. V. Am. J. Hum. Genet. 64, 18-23 (1999).
- 7. Sharma, V. et al. Nature Commun. 9, 1215 (2018).
- 8. Ye, K., Gao, F., Wang, D., Bar-Yosef, O. & Keinan, A. *Nature Ecol. Evol.* **1**, 0167 (2017).
- Goldstein, S., Beka, L., Graf, J. & Klassen, J. Preprint at bioRxiv https://doi.org/10.1101/362673 (2018).

MEDICAL RESEARCH

Immune link to failure of cancer treatment

Prostate-cancer treatment usually fails after time as resistance to therapy develops. It emerges from studies of mice and human cells that a population of immune cells can cause this type of treatment resistance. SEE ARTICLE P.363

MATTHEW D. GALSKY

rostate cancer causes more than 300,000 deaths annually worldwide and is one of the most common causes of cancer-linked mortality in men1. In 1941, the demonstration² that the condition regressed after surgical castration established a link between prostate-cancer growth and androgens — the hormones, such as testosterone, that are mainly generated in the testes and aid the development of male characteristics. The current standard treatment for advanced-stage prostate cancer is androgen depletion by chemical means. However, this almost invariably provides only a temporary halt to the disease. When cancer progression resumes despite low androgen levels, the condition is known as castration-resistant prostate cancer. On page 363, Calcinotto et *al.*³ report that the action of immune cells can drive this type of treatment resistance. This discovery could pave the way to new therapeutic options, and illuminates our understanding of the spectrum of interactions in the prostatecancer microenvironment.

The androgen receptor is a protein that can regulate gene expression. Androgendeprivation therapy can lead to prostate-cancer regression because the absence of androgenmediated signalling causes cancer cells to die or cease dividing^{4,5}. The resulting decrease in tumour size is often monitored in the clinic by tracking a decline in the level of a protein called prostate-specific antigen (PSA). Although it was originally thought that castration-resistant prostate cancer arises through mechanisms that are independent of androgen-mediated signalling, certain observations have challenged that. PSA expression is regulated by the androgen pathway, and an increase in the level of PSA almost always accompanies the development of castration-resistant disease⁶. Moreover, clinical improvement can occur when people undergoing androgen-deprivation therapy are also given extra treatments that hamper androgen

A range of mechanisms underlying castration-resistant prostate cancer have been reported⁸, and several causes probably contribute to disease progression in any given individual. The identification of mechanisms associated with disease progression has led to the development of associated treatments. For example, up to half of castration-resistant prostate cancers⁹ are accompanied by an increase in the number of copies of the

androgen-receptor gene. This, along with other clues suggesting that androgen-receptor expression correlates with disease progression, has stimulated the generation of drugs that can inhibit the androgen receptor¹⁰. Castration resistance can also occur through an increase in the expression of enzymes that synthesize androgens, and this discovery led to the development of drugs that inhibit androgen biosynthesis¹¹. However, resistance to both these types of inhibitor can develop, so there is still a need for additional clinical strategies.

Calcinotto and colleagues investigated whether immune cells might aid the development of castration-resistant prostate cancer. They focused on an immune-cell population known as myeloid-derived suppressor cells (MDSCs), which includes monocytes and neutrophils that might be in an immature state of abnormal activation. MDSC presence is linked to poor prognosis for patients who have prostate cancer¹², although this connection has been attributed to the suppression of an anticancer immune response. Calcinotto et al. observed that tumour biopsies from people who have developed castration-resistant prostate cancer contain more MDSCs that express the proteins CD11b, CD33 and CD15 than do samples from people whose prostate cancer has not progressed to the castration-resistant stage.

The authors investigated whether MDSCs might contribute directly to castration resistance, using human cell samples and mouse models of prostate cancer. In mice, the authors found that surgical castration resulted in an increase in the recruitment of MDSCs to tumours, compared with the recruitment observed in control animals given a mock operation. Calcinotto and colleagues grew mouse MDSCs *in vitro* and isolated samples of the culture medium. When this medium was added to androgen-dependent

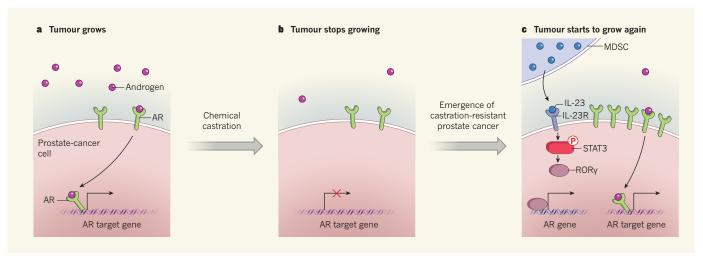


Figure 1 | **An immune cell drives treatment resistance in prostate cancer.** Calcinotto *et al.*³ used mouse models and human clinical samples to investigate how prostate cancer evades therapy. **a**, If male hormones called androgens bind to the androgen receptor (AR), this can drive the expression of genes that promote prostate-cancer growth. **b**, A standard treatment for the disease is chemical castration, in which drugs are used to decrease androgen levels. However, the subsequent slowing of tumour progression is not permanent. **c**, When tumour growth returns, the condition is called

castration-resistant prostate cancer. Calcinotto and colleagues found that a type of immune cell called a myeloid-derived suppressor cell (MDSC) can cause this treatment failure. If an MDSC cell secretes a protein called IL-23, this might bind to the IL-23 receptor (IL-23R) on tumour cells. This binding triggers a pathway in the tumour cell mediated by the proteins ROR γ and STAT3 (the latter is phosphorylated; P is a phosphate group), which can drive AR expression. This increase in AR expression helps to drive the androgen-dependent gene expression that boosts prostate-cancer growth.

prostate-cancer cell lines cultured *in vitro* under androgen deprivation, it sustained the proliferation and survival of the cells and caused an increase in the transcription of genes whose expression is driven by the androgen receptor.

The authors carried out equivalent experiments using human cells and made similar findings. Furthermore, the use of pharmacological techniques to deplete MDSCs delayed the emergence of castration resistance in mice. Together, these results suggest that MDSCs secrete a factor that drives the emergence of castration-resistant prostate cancer.

To identify this key factor, the authors took samples of tumours and associated cells from castrated mice and from animals that underwent a mock operation, and searched for the genes that showed the greatest increase in expression in the samples from castrated mice. Their results included the gene encoding IL-23 (an immune signalling protein called a cytokine) and a gene that encodes a subunit of the receptor to which IL-23 binds. Analysis of prostate-cancer specimens from the clinic confirmed the importance of IL-23, and there were more IL-23-expressing MDSCs in castration-resistant prostate-cancer specimens than in specimens from tumours that were not castration resistant.

Calcinotto and colleagues propose that signalling mediated by MDSC-secreted IL-23 and by the IL-23 receptor on prostate-cancer cells promotes the development of castration-resistant prostate cancer. Using pharmacological or genetic approaches to block IL-23-mediated signalling in mice, they obtained evidence that such treatment delays the development of castration-resistant prostate cancer.

The authors carried out studies to determine

the signalling pathway downstream of IL-23 that mediates the return of tumour growth, and focused on two IL-23-regulated proteins (STAT3 and RORy) that are part of a pathway that boosts androgen-receptor signalling¹³. Their results are consistent with a model in which IL-23-mediated activation of the STAT3-RORy pathway leads to an increase in expression of the androgen receptor and an increase in expression of genes whose transcription depends on that receptor (Fig. 1). Strikingly, the authors demonstrated that if mice that had developed castration-resistant prostate cancer were given an antibody that blocks IL-23 and an androgen-receptor inhibitor called enzalutamide, this reversed castration resistance and caused the animals' tumours to shrink.

Calcinotto and colleagues' work has important clinical implications and advances our understanding of the biological processes that underlie castration resistance. Antibodies that block IL-23 are approved for clinical use to treat autoimmune conditions, which clears the way for them to be tested as a possible treatment for castration-resistant prostate cancer. The findings also raise the question of whether immune cells might contribute to the progression of other sorts of cancer in which growth is driven by hormone-receptor signalling. Some controversy currently exists about whether MDSCs are a cell population that is distinct from normal neutrophils and monocytes, given that MDSCs are highly similar to those cells, yet are considered functionally different. There is not yet a clear consensus about how to identify MDSCs on the basis of the expression of cell-surface markers¹². This issue could affect future studies of these cells.

As with any disease mechanism studied

mainly in animal models, the prevalence of possible MDSC-associated castration resistance in humans remains to be determined. It might be a minor mechanism in most patients, a major mechanism in a minority of patients or somewhere in between. The scale of the effect of MDSCs, and the ability to select the specific people likely to respond to treatment targeting castration-resistant tumours, will probably be crucial in determining whether such therapy against prostate cancer is successfully implemented in the clinic.

Matthew D. Galsky is in the Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, and at the Tisch Cancer Institute, New York, New York 10029, USA.

e-mail: matthew.galsky@mssm.edu

- Torre, L. A. et al. CA Cancer J. Clin. 65, 87–108 (2015).
- Huggins, C. & Hodges, C. V. J. Urol. 168, 9–12 (2002).
- 3. Calcinotto, A. et al. Nature **559**, 363–369 (2018).
- 4. Pernicová, Z. et al. Neoplasia **13**, 526–536 (2011).
- Westin, P., Stattin, P., Damber, J. E. & Bergh, A. Am. J. Pathol. 146, 1368–1375 (1995).
- Scher, H. I. et al. J. Clin. Oncol. 26, 1148–1159 (2008).
- Ryan, C. J. & Small, E. J. Curr. Oncol. Rep. 7, 228–233 (2005).
- Seruga, B., Ocana, A. & Tannock, I. F. Nature Rev. Clin. Oncol. 8, 12–23 (2011).
- 9. Grasso, C. S. et al. Nature **487**, 239–243 (2012). 10. Tran, C. et al. Science **324**, 787–790 (2009).
- 11. Attard, G. et al. J. Clin. Oncol. **26**, 4563–4571 (2008).
- 12. Veglia, F., Perego, M. & Gabrilovich, D. *Nature Immunol.* **19**, 108–119 (2018).
- 13.de Miguel, F., Lee, S. O., Onate, S. A. & Gao, A. C. *Nucl. Recept.* **1**, 3 (2003).

The author declares competing financial interests. See go.nature.com/2lhcq9t for details.

This article was published online on 27 June 2018.

CLIMATE SCIENCE

Atlantic slowdown boosts surface warming

The circulation system of the North Atlantic Ocean has weakened and is predicted to weaken further in the coming decades. An analysis suggests that this decline could lead to accelerated global surface warming. SEE LETTER P.387

GERARD D. MCCARTHY & PETER W. THORNE

from 1975 to 1998, but this growth then slowed somewhat for about 15 years — an event that gained popular attention 1 as a 'hiatus'. Since then, we have experienced the four warmest years on record, which has served to dampen popular interest in the event. However, because climate change is a complex response to slowly varying external drivers, it is important to fully understand past climate behaviour and the underlying causes. On page 387, Chen and Tung' report that the system of ocean currents known as the Atlantic Meridional Overturning Circulation (AMOC)

can explain changes in rates of global surface warming. Rather than the conventional picture of a vigorous AMOC associated with elevated surface temperatures in the Atlantic Ocean, the authors emphasize the role of the AMOC in taking heat from the surface and storing it in the deep ocean.

The connection between the AMOC and variations in the heat content of the subpolar North Atlantic Ocean has long been acknowledged. The AMOC transports heat northwards to the subpolar North Atlantic and to the Greenland, Iceland and Norwegian Seas. There, through a range of processes, deep water is formed that moves as a southward cold flow. This conveyor belt of northward-flowing, warm, shallow water and southward-flowing, cold, deep water defines the AMOC.

Relative to latitudinal averages, surface temperatures could be 5 °C cooler in the subpolar North Atlantic Ocean and up to 10 °C cooler in the Norwegian Sea if the AMOC were absent³. Consequently, a strong AMOC is typically associated with warming in the Northern Hemisphere. This association is consistent with evidence from palaeoclimatology that suggests that, during the most recent ice age, warmer periods coincided with a

vigorous AMOC and colder periods coincided with a weak AMOC⁴.

Chen and Tung's study emphasizes a different role for the AMOC in the modern climate. Atmospheric concentrations of greenhouse gases are currently being increased at a rate that is unprecedented in millennia and most likely millions of years. As a result, the role that climate mechanisms might have had in the past might not be a good guide to their current or future role. The authors contend that half of the heat arising from ever-increasing greenhouse-gas concentrations is stored in the deep waters of the North Atlantic when the AMOC is increasing, thereby reducing overall global surface warming (Fig. 1).

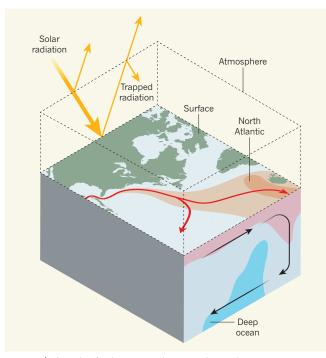


Figure 1 | The role of Atlantic circulation in the modern climate. Increasing atmospheric concentrations of greenhouse gases mean that more incoming solar radiation is trapped in the atmosphere, a consequence of which is surface warming. Chen and Tung² report that the ocean circulation system known as the Atlantic Meridional Overturning Circulation (AMOC) can offset this warming. The AMOC transports heat northwards (red arrows) in the North Atlantic Ocean; the light- and darkorange colours indicate regions that are 5 °C and 10 °C warmer, respectively, than latitudinal averages. The authors emphasize the role of the AMOC in transporting heat from the surface to the deep ocean (black arrows); the pink and blue colours indicate regions where heat increases and decreases, respectively, when the AMOC is increasing (based on Fig. 2a of ref. 2).

The authors show that a cycle of increasing and then decreasing AMOC from the 1940s to the mid-1970s coincided with a period of global-warming slowdown; a quiescent period of weak AMOC from the mid-1970s to the late 1990s coincided with rapid global warming; and an increase in AMOC strength from the late 1990s to 2005 and a decrease thereafter coincided with the 'hiatus' in global warming (see Fig. 3 of the paper²).

When the causes of the 'hiatus' were first being investigated, the Atlantic was not an obvious place to look. The focus was on the Pacific Ocean because the tropical Pacific was one of the only places where surface temperatures did not rise during this period⁵. Understanding of the event developed as several factors were taken into account, including the effect of changes in ocean heat content across multiple ocean basins⁶. Chen and Tung now bring focus to the North Atlantic. Their work suggests that the warm surface temperatures there were indicative of an increasing AMOC and that the associated increase in ocean heat uptake played a key part in the 'hiatus'.

One of the main caveats of Chen and Tung's study is that, by necessity, the authors used proxies for AMOC strength because no direct observations of sufficient length exist. There are only four observatories that measure the

AMOC across the full width of the Atlantic: SAMBA at 34.5°S, RAPID at 26°N, NOAC at 47°N and OSNAP between 53°N and 60°N. The longest-running, RAPID, was deployed in only 2004. These observatories need to be maintained for many decades if we are to fully understand the role of the AMOC in our changing climate.

There is much to be done to determine how the AMOC affects surface temperature in different regions and on different timescales. For instance, Chen and Tung highlight the potential role of the Southern Ocean in heat uptake in the period since 2005. Such a feature could be part of a see-saw pattern of alternating heat uptake by the North Atlantic and Southern Ocean.

There is also a distinct difference between the effects of decadal AMOC variability and of an AMOC collapse on global temperatures. Although the prospect of the AMOC passing a tipping point and collapsing is considered unlikely, it is not impossible, and an event this dramatic could lead to global surface cooling⁷. The threshold between a weak AMOC that reduces ocean heat uptake, allowing global surface temperatures to rise unabated, and a very weak or collapsed AMOC that

causes cooling in the North Atlantic and global surface warming to slow or stop will be a key point of debate.

The AMOC is deemed "very likely" to weaken in the coming decades¹. Indeed, the Atlantic has seen muted rises in surface temperature relative to the global ocean over the past few decades. This relative lack of warming has been interpreted as a fingerprint of AMOC decline, potentially linked to anthropogenic climate change⁸. Whether the AMOC observatories will document the predicted decline remains to be seen, but they have already observed that the AMOC is in a weakened state⁹. Chen and Tung predict that such a weak AMOC will result in a period of rapid global surface warming that could last for more than two decades. ■

Gerard D. McCarthy and Peter W. Thorne

are at the Irish Climate Analysis and Research Units (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland

e-mails: gerard.mccarthy@mu.ie; peter.thorne@mu.ie

- Intergovernmental Panel on Climate Change (IPCC). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2014).
- 2. Chen, X. & Tung, K.-K. Nature **559**, 387–391 (2018).
- 3. Jackson, L. C. et al. Clim. Dyn. **45**, 3299–3316 (2015).
- 4. Kosaka, Y. & Xie, S.-P. *Nature* **501**, 403–407 (2013).
- Medhaug, I., Stolpe, M. B., Fischer, E. M., & Knutti, R. Nature 545, 41–47 (2017).
- Broecker, W. S. Oceanography 4, 79–89 (1991).
 Drijfhout, S. Sci. Rep. 5, 14877 (2015).
- Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. Nature **556**, 191–196 (2018).
- Smeed, D. A. et al. Geophys. Res. Lett. 45, 1527–1533 (2018).

BIODIVERSITY

Speciation far from the madding crowd

New species of marine fishes are found to emerge at a faster rate in high-latitude oceans, which have lower densities of species, than in the species-rich tropics. Are the tropics too crowded for new species to take hold? See Letter P.392

ARNE O. MOOERS & DAN A. GREENBERG

The tropics are, like many cities, hot, busy and crowded. It was previously thought¹ that these conditions in the tropics generate a hotbed for the formation of new species (speciation). Species diversity is remarkably high in the tropics and declines toward the poles. However, newly developed tools to measure speciation rates, coupled with ever-growing global data sets, have enabled the surprising finding that terrestrial speciation rates for the past few million years are similar across different latitudes² or increase outside the tropics³. On page 392, Rabosky et al.4 document a speciation rate for marine fishes at high latitudes that is twice the speciation rate in tropical seas. This high speciation rate in cold, species-poor waters poses an interesting conundrum for evolutionary biologists and ecologists.

There are two potential drivers of high speciation rates in the tropics. First, the elevated temperatures in the region both speed up metabolism, increasing the number of mutations, and decrease generation times. This is a potentially powerful combination, producing more of the variation necessary for evolution and the possibility of faster evolution. A second possible driver is ecological opportunity. The energy-rich tropics offer abundant resources that can support many different niches. And the

tropics are so rich in species that the interactions of members of a single species with its competitors, predators and parasites might differ from place to place, leading to different adaptations and eventual divergence into new niches¹. Although this narrative makes for a compelling theory, Rabosky and colleagues' discovery suggests a different story, at least for marine fishes.

The authors gathered genetic data for 11,638 species of marine and freshwater fish, along with information on inferred evolutionary relationships based on taxonomic groupings for 19,888 additional fish species for which genetic data were not available. Using these data, and information from 139 dated fossil fishes, the authors generated a large set of plausible phylogenetic trees detailing the evolutionary relationships between all living marine fishes, and, crucially, estimates of when different lineages diverged from one another. These dated trees enable speciation rates to be inferred on the basis of the branching patterns of the tree. Species connected by short branches, and with many close relatives, have high speciation rates, whereas species that are separated by long branches and that have few close relatives have low speciation rates.

Most taxonomic groups are made up of lineages with both low and high speciation rates. The marine fishes in the authors' large phylogenetic trees were no exception, with



50 Years Ago

More than 7,000 people died in traffic accidents last year in Britain and nearly 94,000 were seriously injured ... The causes and possible preventative measures were the subject of a recent conference ... A police traffic superintendent is reported to have said that the defect which causes nearly every accident lies in "the nut holding the wheel". One contributor pointed out that road engineers should not assume that drivers are omniscient. "If asked to make more than one decision at a time, they will fail; if faced with a situation which can be misinterpreted, someone will eventually find the wrong meaning". The effect of human fallibility is easily apparent in the statistics ... during the last three months of 1967 ... after the breathalyser test came into operation ... driver and passenger casualties fell by 19 per cent, motor cyclist casualties by 16 per cent and pedal cyclists by 14 per cent. From Nature 20 July 1968

100 Years Ago

The utility of forests to a nation is one of the economic factors to its well-being which have been brought to an unforeseen prominence during the world-war: and perhaps to no other European nation has this ... development proved so startling, because so totally unsuspected, as to ourselves. Our woods were not grown from the commercial aspect — sport, amenity, and shelter to crops and stock were their main raison d'être. We did not consider it necessary to grow woods for purely commercial reasons — that is, for the sake of the timber and pit wood and paper pulp, etc ... We obtained our requirements in these commodities by importing them from abroad, and relied on the Navy being able to safeguard these imports. We have now discovered our mistake and are paying for it. From Nature 18 July 1918



Figure 1 | Remarkable fish diversification in remarkably cold waters. Rabosky et al. 4 analysed global patterns of where new species of marine fishes arise. The authors find that marine fishes that live near Earth's poles, such as Antarctic icefishes (for example, the black-fin icefish, Chaenocephalus aceratus, shown here), give rise to new species much faster than do fishes in warmer and more species-rich tropical seas.

speciation rates varying by more than 50-fold between lineages. The authors combined these values with global maps of where these species live, revealing a clear geographical structure to the speciation rates. Because small biases integrated over large amounts of data could produce misleading inferences in these sorts of studies⁵, the authors considered several ways to estimate speciation rates and to map species' location. But whether the authors considered the patterns looking species by species, place by place or ecoregion by ecoregion, there was always a pattern of the average speciation rate increasing from the tropics towards the poles.

Rabosky and colleagues consider only a few potential mechanisms that might affect the rate at which marine fishes produce new species at high latitudes. At higher latitudes, marine fishes tend to have longer generation times and slower metabolisms⁶ than have fishes in the tropics, suggesting that such extended generation times and lower mutational input do not limit speciation rate in these cold-adapted lineages. The authors also tested and discounted the interesting possibility that high-latitude species are the descendants of tropical lineages that exhibited adaptations for cold-water living and also happened to have high speciation rates. This negative result suggests that a high-latitude marine environment, rather than the species that colonizes it, drives high speciation rates.

The correlation of high speciation rates with low diversity is consistent with the idea that there are unfilled ecological opportunities near the poles. However, ecological opportunity is something that is inferred rather than witnessed¹. The shape of a phylogenetic tree can indicate slowing speciation as species numbers rise — a pattern that is consistent with diminishing ecological opportunity⁷. If closely related species in such clades occupy different niches, as might be the case for the high-latitude Antarctic icefishes8 (Fig. 1), this

would be consistent with ecological opportunity having a key role in driving their diversification. Such analyses are needed to determine whether high-latitude groups have reached the ecological limits of their ecosystems or whether high-latitude fish diversity might be expected to continue increasing.

Continued diversification at higher latitudes might seem reasonable, given that Earth's cooling over the past 30 million years or so⁹ has given rise to the present, relatively young temperate and polar realms. However, the rate of speciation at high latitudes reported by Rabosky (roughly 0.2 new species per species per million years) is high. If this rate had been sustained over the whole of the past 30 million years, high latitudes would have tropical levels of species diversity by now. Given that the accumulation of diversity depends on both speciation and extinction rates, one explanation that reconciles a high speciation rate and low current diversity is if both speciation and extinction are elevated outside the tropics¹⁰. This could result in a pattern in which an increase in the number of species is limited by high extinction rates, and poleward realms would have few, but relatively young, species. Measuring extinction rates is almost as difficult as trying to assess ecological opportunity, but new approaches that combine information on extinct species represented in the fossil record with information from their living relatives¹¹ might offer a way to investigate whether extinction rates are greater at higher latitudes.

The view of high-latitude oceans as 'sleepy' backwaters remote from the exciting evolutionary bustle of the tropics will need to change if it turns out that both speciation and extinction of marine fishes occur at a faster pace beyond the tropics. Such a pattern would imply that living cheek by jowl, or rather gill by jaw, in the tropics is a condition that is more constraining than productive, such that the real biodiversity action is taking place where there is less, rather than more, biodiversity. Far from the madding crowd, as it were.

Arne O. Mooers and Dan A. Greenberg are in the Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. e-mails: amooers@sfu.ca; dgreenbe@sfu.ca

- 1. Schluter, D. Am. Nat. 187, 1-18 (2016).
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. Nature 491, 444-448 (2012).
- Schluter, D. & Pennell, M. W. Nature **546**, 48–55
- 4. Rabosky, D. L. et al. Nature 559, 392-395 (2018). Ricklefs, R. & Pagel, M. Nature 491, 336–337
- (2012).6. Charnov, E. L. & Gillooly, J. R. Integr. Comp. Biol. 44,
- 494-497 (2004) 7. Etienne, R. S. et al. Proc. R. Soc. B 279, 1300-1309
- (2012). 8. Near, T. J. et al. Proc. Natl Acad. Sci. USA **109**, 3434-3439 (2012).
- Zachos, J., Paganini, M., Sloan, L, Thomas, E. & Billups, K. Science 292, 686-693 (2001).
- 10. Weir, J. T. & Schluter, D. Science 315, 1574-1576
- 11. Mitchéll, J. S., Etienne, R. S. & Rabosky, D. L. Syst. Biol. https://doi.org/10.1093/sysbio/syy035 (2018).

This article was published online on 4 July 2018.

CORRECTION

The News & Views article 'Intestinalniche conundrum solved' (Nature 558, 380-381; 2018) indicated that two papers (M. Shoshkes-Carmel et al. Nature 557, 242-246; 2018, and B. Degirmenci et al. Nature 558, 449-253; 2018) solved an outstanding debate — the identify of a stromal-cell population that sends Wnt signals to intestinal stem cells. However, a paper published earlier this year (G. Greicius et al. Proc. Natl Acad. Sci. USA 115, E3173-E3181; 2018) also identified a stromal-cell source for Wnt signals.



Electron ptychography of 2D materials to deep sub-ångström resolution

Yi Jiang^{1,6}, Zhen Chen^{2,6}, Yimo Han², Pratiti Deb^{1,2}, Hui Gao^{3,4}, Saien Xie^{2,3}, Prafull Purohit¹, Mark W. Tate¹, Jiwoong Park³, Sol M. Gruner^{1,5}, Veit Elser¹ & David A. Muller^{2,5}*

Aberration-corrected optics have made electron microscopy at atomic resolution a widespread and often essential tool for characterizing nanoscale structures. Image resolution has traditionally been improved by increasing the numerical aperture of the lens (α) and the beam energy, with the state-of-the-art at 300 kiloelectronvolts just entering the deep sub-ångström (that is, less than 0.5 ångström) regime. Two-dimensional (2D) materials are imaged at lower beam energies to avoid displacement damage from large momenta transfers, limiting spatial resolution to about 1 ångström. Here, by combining an electron microscope pixel-array detector with the dynamic range necessary to record the complete distribution of transmitted electrons and full-field ptychography to recover phase information from the full phase space, we increase the spatial resolution well beyond the traditional numerical-aperture-limited resolution. At a beam energy of 80 kiloelectronvolts, our ptychographic reconstruction improves the image contrast of single-atom defects in MoS₂ substantially, reaching an information limit close to 5α , which corresponds to an Abbe diffraction-limited resolution of 0.39 ångström, at the electron dose and imaging conditions for which conventional imaging methods reach only 0.98 ångström.

The ability to image individual atoms is essential for characterizing structure and defects in 2D materials^{1–3}. In scanning transmission electron microscopy (STEM), the most common technique for achieving

atomic resolution is high-angle annular dark-field (ADF) imaging, which records electrons scattered through large angles to form an incoherent image. The maximum spatial information contained in

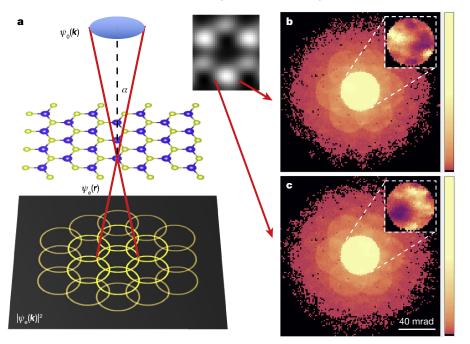


Fig. 1 | **STEM imaging using the EMPAD. a**, At each scan position, the incident probe $(\psi_0(k))$ is focused on the sample and the entire diffraction pattern of the exit wave $(|\psi_e(k)|^2)$ is recorded by the EMPAD. The blue and yellow atoms represent molybdenum and sulfur atoms in the object plane. ψ_0 and ψ_e refer to the incident and exit wavefunctions respectively; \boldsymbol{r} is the (x,y) positional coordinate in the real-space plane; and \boldsymbol{k} is the (k_x,k_y) wavenumber coordinate in the conjugate momentum-space plane.

b, **c**, Averaged diffraction pattern intensity (on a logarithmic scale) from the electron beam at the marked scan positions near a molybdenum column. Insets show the intensity (on a linear scale) of the bright-field disks. The substantial intensity differences at large scattering angles provide contrast information for ADF imaging and are essential for resolution enhancement in ptychography.

¹Department of Physics, Cornell University, Ithaca, NY, USA. ²School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA. ³Department of Chemistry, Institute for Molecular Engineering, James Franck Institute, University of Chicago, Chicago, IL, USA. ⁴Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA. ⁵Kavli Institute at Cornell for Nanoscale Science, Ithaca, NY, USA. ⁶These authors contributed equally: Yi Jiang, Zhen Chen. *e-mail: david.a.muller@cornell.edu

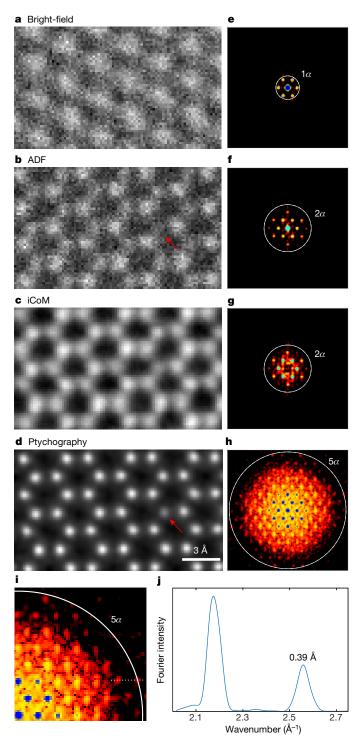


Fig. 2 | Comparison of different imaging techniques using 4D EMPAD dataset measured from monolayer MoS₂. a, Coherent bright-field image. b, Incoherent ADF image. c, iCoM image. d, Phase of the transmission function reconstructed using full-field ptychography. The red arrows indicate a sulfur monovacancy that is detectable in ptychography. e-h, False-colour diffractogram intensities (on a logarithmic scale) of the bright-field (e), ADF (f), iCoM (g) and full-field ptychography (h) images. The information limit (white circle) of ptychography is close to 5α (107 mrad); the information limits of the other imaging methods are also shown. i, Close-up of h. j, Line profile along the dotted horizontal white line in i (linear scale) across two diffraction spots. The peak at 5α corresponds to an Abbe resolution of 0.39 Å.

an ADF image (or other incoherent imaging modes) is determined by the momentum transfer across the diameter of the probe-forming aperture—that is, twice the semi-convergence angle (α)^{4,5}. Therefore,

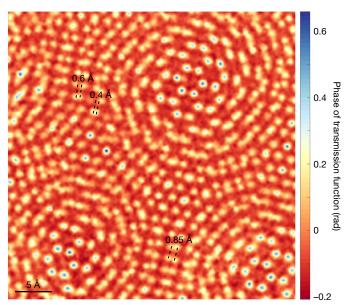


Fig. 3 | Real-space resolution test of full-field ptychography using twisted bilayer MoS₂. The two sheets are rotated by 6.8° with respect to each other, and the misregistration of the molybdenum atoms provides a range of projected distances that vary from a full bond length down to complete overlap. Atoms are still cleanly resolved at a separation of 0.85 ± 0.02 Å, with a small dip still present between atoms separated by about 0.61 ± 0.02 Å, similar to the contrast expected for the Rayleigh criterion for conventional imaging. Atom-pair peaks at 0.42 ± 0.02 Å show a 6% dip at the midpoint, suggesting that the Sparrow limit lies just below 0.4 Å. The Raleigh resolution for ADF STEM is 1.2 Å for these imaging conditions (Extended Data Fig. 3a).

obtaining high-resolution images generally requires small wavelengths and large apertures, and the latter in turn introduces phase-distorting artefacts from geometrical and chromatic aberrations. The demonstration of practical aberration correctors^{6,7} has ameliorated these phase errors substantially, and for the past decade the state-of-the-art for ADF images has reached the deep sub-angström regime of about 0.5-Å resolution at 300 keV^{8,9}, which is sufficient for imaging most bulk materials. On the other hand, the characterization of 2D materials, such as single-defect detection and imaging of interface or edge structures, always requires lower beam energies (roughly 20-80 keV) to minimize knock-on damage^{2,10,11}. Because lower energies imply longer electron wavelengths, the resolution of ADF imaging is reduced substantially and reaching sub-angström resolution is possible only with specialized correctors that correct both geometric and chromatic aberrations or with monochromatic electron beams ^{12,13}. Moreover, ionization damage, which cannot be avoided by lowering the beam voltage, also restricts the electron dose applied to the sample, limiting the ultimately achievable signal-to-noise ratio¹⁴, further reducing image resolution and contrast.

However, it has long been recognized that the information limit set by diffractive optics is not an ultimate limit¹⁵. There is phase information encoded throughout a diffraction pattern formed from a localized electron beam, in the form of interference patterns between overlapping scattered beams (Fig. 1a). As the incident localized beam is scanned, this phase information and hence the interference patterns change in a predictable manner that can be used to retrieve the phase differences—an approach known as ptychography^{16–18}. Although originally conceived to solve the phase problem in crystallography, modern ptychography is equally applicable to non-crystalline structures^{19–22} and has received renewed attention as a dose-efficient technique^{23,24} for recovering the projected potential of thin materials, with modifications for measuring finite-thickness and three-dimensional samples^{25,26}. In principle, the resolution is limited by the largest scattering angle at which meaningful information can still be recorded; however, because

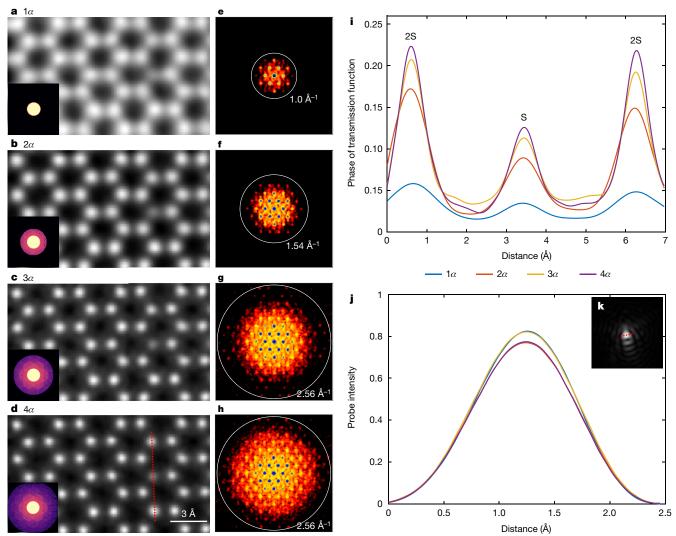


Fig. 4 | Ptychographic reconstructions using data with different cutoff angles. \mathbf{a} - \mathbf{d} , Ptychographic reconstructions using electrons collected using cutoffs of 1–4 times the aperture size (α) . The averaged diffraction patterns are shown in the lower-left corner of each image. \mathbf{e} - \mathbf{h} , False-colour diffractograms (on a logarithmic scale) of the reconstructions in \mathbf{a} - \mathbf{d} . The white circles indicate the information limit. \mathbf{i} , Line profiles

across three sulfur columns (indicated by the vertical red dashed line in **d**). The columns containing two sulfur atoms are labelled 2S; the column with one sulfur atom is labelled S. **j**, Line profiles across the reconstructed probe function (dashed line in **k**) at different cutoffs. **k**, Probe profile reconstructed by ptychography using the dataset with a 4α cutoff.

electron-scattering form factors have a very strong angular dependence, the signal falls rapidly with scattering angle, so a detector with high dynamic range and sensitivity is required to exploit this information.

Ptychography has been widely adopted for light²⁷ and X-ray^{19,28} applications, but the technique is still underexplored in transmission electron microscopy, in large part because of the detector challenges. Traditional electron cameras such as charge-coupled devices (CCDs) and pixelated detectors have been hampered by slow readout speed or poor dynamic range. Previous work^{23,29–35} has mainly made use of electrons only within the bright-field disk; therefore, the image resolution did not overcome the 2α limit imposed by the physical aperture. The first attempt¹⁷ at demonstrating super-resolution ptychography involved phasing the Fourier coefficients of silicon out to the (400) reflection to reconstruct the unit cell with a resolution of 1.36 Å. However, this result determined only structure factors, limiting its application to periodic crystalline structures. A more recent demonstration³⁶ for a lower-resolution scanning electron microscope equipped with a CCD camera showed that the resolution of iterative ptychographic reconstructions can be improved when using information at higher scattering angles.

There are three challenges to improving resolution and dose efficiency to the point needed to advance beyond the current state-of-the-art diffractive imaging. First, a detector must be able to record the full range of scattered intensities without introducing nonlinear distortions or saturating the central beam. Second, the detector must not only possess single-electron sensitivity, but also retain a high detective quantum efficiency when summing over the large ranges of empty pixels at high scattering angles. Third, each diffraction pattern must be recorded rapidly enough that the full image is not sensitive to drift and instabilities in the microscope, which usually leaves only a few minutes to record a full four-dimensional (4D) dataset. The combination of the first and third conditions poses an additional constraint that the detector must also have a high dynamic current range. Therefore, it is not sufficient to count single electrons for a long time at a low beam current; instead, large currents per pixel need to be recorded in very short times. Most pulse-counting methods are limited to about 2–10 MHz by the transit time of the electron cloud through the silicon detector. This translates to 0.3–1.6 pA per pixel, although few systems reach this limit. Instead, to keep nonlinearities below 10%, a limit of 0.03 pA per pixel is more typical³⁷. Direct charge integration in a CCD geometry is even more limited by the well depth, to about 20 electrons per pixel per frame. At a 1-kHz frame, this corresponds to 0.003 pA per pixel—a limit at which single-frame Poisson statistics would then be below the Rose criterion for contrast detectability³⁸.

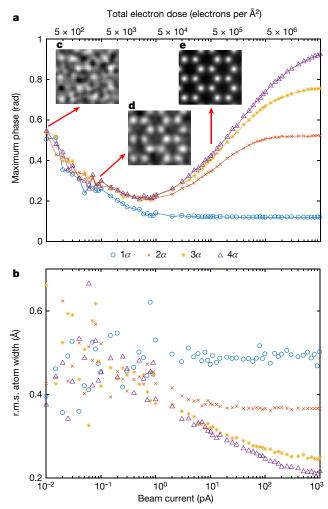


Fig. 5 | Simulation study of full-field ptychography as a function of cutoff angle and beam current. a, b, Reconstruction resolution evaluated by the maximum range of the reconstructed phase (a) and the root-mean-square (r.m.s.) width of the molybdenum atom defined by the standard deviation of a fitted Gaussian (b). At large electron doses, the resolution is determined by the maximum detector angle. As current decreases, the resolution is instead limited by the Poisson noise. c-e, Reconstructed phase maps using diffraction patterns with 4α cutoff at beam currents of 0.01 pA (c), 0.1 pA (d) and 10 pA (e), as indicated by the arrows. The increase in the phase range and the decrease of the r.m.s. atom width at large dose (more than 1 pA) is a measure of the resolution improvement, but the increase in the phase range at low dose (less than 1 pA) reflects increasingly large noise fluctuations in the reconstructions, which are also evident in the increased variations of the r.m.s. atom width.

To overcome these challenges, we developed an electron microscope pixel-array detector (EMPAD)³⁹ that is capable of recording all the transmitted electrons with sufficient sensitivity and speed to provide a complete ptychographic reconstruction. Our EMPAD design has a high dynamic range of 1,000,000:1 while preserving single-electron sensitivity with a signal-to-noise ratio of 140 for a single electron at 200 keV³⁹. The detector retains a good performance from 20 keV to 300 keV. Here we operate at 80 keV, at which the signal-to-noise ratio per pixel is 50 for a single electron, the detective quantum efficiency is 0.96 and the maximum beam current per pixel is 5 pA. By using essentially all of the electrons collected (99.95% of the transmitted beam, as determined using multi-slice simulations), with a full 4D dataset acquired in typically a minute, our full-field ptychographic reconstructions roughly double the image resolution compared to the highest-resolution conventional single-channel imaging modes, such integrated centre-of-mass $(iCoM)^{40,41}$ and ADF STEM.

Data acquisition and reconstruction

In Fig. 1a we show a schematic of the experimental configuration with the EMPAD. To minimize radiation damage, a monolayer of MoS₂ is imaged at a primary beam energy of 80 keV. At each (x, y) scan position, the EMPAD records a diffraction pattern (k_x, k_y) from the convergent probe, thus forming a 4D dataset (x, y, k_x, k_y) . In Fig. 1b, c we show averaged diffraction patterns corresponding to two positions near a single molybdenum column. Supplementary Videos 1 and 2 show continuous evolutions of averaged and raw diffraction patterns, respectively, at various scan positions, where, compared to Fig. 1b, c it is easier to observe the intensity variations in the overlaps between the higher-order diffraction disks that occur as the relative phase of the interfering beams changes with position. The considerable changes in the distribution outside the central disk provides essential contrast information in ADF images; the resolution improvement in full-field ptychography over previous bright-field ptychographic methods stems from exploiting the phase information encoded in the contrast between overlapping higher-order disks. The angle-averaged radial distribution function of the position-averaged diffraction pattern (Extended Data Fig. 1) shows a four-orders-of-magnitude intensity range for our dataset. The reconstruction algorithm is implemented using the extended ptychographic iterative engine (ePIE) algorithm^{20,42}, which reconstructs the transmission function iteratively and refines the probe function to accommodate aberrations and noise. We also compare our performance to the simpler Wigner-distribution deconvolution (WDD)⁴³, which in its simplest form assumes a known probe function and uses the information within the central disk, and shows a similar performance to using ePIE on only the central disk (that is, with a cutoff α) or iCoM imaging. In principle, WDD could also utilize the dark-field signal and surpass the aperture-limited resolution if a proper de-noising strategy is applied^{43,44}.

The 4D EMPAD data can generate all elastic imaging modes for benchmarking from the same dataset, including coherent bright-field, iCoM and ADF modes. As shown in Fig. 2a, e, the coherent bright-field image has the poorest resolution (restricted to within α , as expected). The incoherent ADF image (Fig. 2b, f) doubles the information limit (from α to 2α), but is limited by a low signal-to-noise ratio and residual probe aberrations. Although the iCoM image is less noisy, its resolution is still within 2α (Fig. 2g) because the structural information is influenced by the incident probe via convolution. By contrast, full-field ptychography recovers the phase of the transmission function directly (Fig. 2d) and achieves an information limit of 5α (Fig. 2h). Noise artefacts are also reduced substantially and the light-atom sulfur monovacancy (indicated by red arrows) is resolved more clearly. In Fig. 2i, j we show an enlarged section of the Fourier intensity map from the ptychographic reconstruction and a line profile across a diffraction spot at the 5α limit, demonstrating an estimated Abbe resolution⁴⁵ of 0.39 Å or better (there are higher-order spots of weaker intensity but they are not as uniform in all directions). For comparison, with our electron optical conditions, the expected Abbe resolution for conventional incoherent imaging modes such as ADF STEM is 2α or 0.98 Å.

A second measure of spatial resolution is the minimum resolvable distance between two atoms. For 2D materials, this measure is complicated by the fact that it requires atoms to be spaced closer than the shortest known bond lengths. To accomplish this test, we use a twisted bilayer sample of two MoS₂ sheets rotated by 6.8° with respect to each other. This effectively creates an incommensurate atomic moiré pattern, which provides projected Mo-Mo atomic distances that vary from a full bond length to fully superimposed atoms, with many intermediate distances across the incommensurate moiré quasi-periodicity of 28 Å (see, for example, Fig. 1c of ref. 46). In Fig. 3 we show the ptychographic reconstruction across a moiré supercell, in which atomic columns midway between the aligned regions are resolved as separate atoms at 0.85 ± 0.02 Å. The dominant uncertainty here, and in our other distance measurements, is the systematic error from scan distortions rather than random errors from counting statistics. The dip between adjacent columns can still be seen at 0.60 ± 0.02 Å—close to the Rayleigh limit

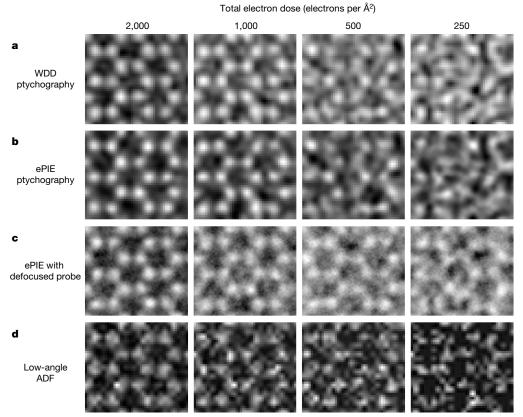


Fig. 6 | Comparison between ptychographic techniques and low-angle ADF imaging at low electron doses. a, b, Ptychographic reconstructions of simulated data with an in-focused probe, using the WDD (a) and ePIE (b) methods. c, ePIE reconstructions of data with a large defocused probe. d, Low-angle ADF image (integrating from 1α to 4α) using the same

simulated dataset as for \mathbf{a} and \mathbf{b} . The defocused ePIE approach (\mathbf{c}) shows a dose advantage over the other two ptychography approaches (\mathbf{a},\mathbf{b}) by a factor of roughly two or better, and over low-angle ADF imaging (\mathbf{d}) —currently the optimal single-channel imaging method for 2D materials—by a factor of roughly four.

for resolution. Atom-pair peaks measured at 0.42 ± 0.02 Å show a 6% dip at the midpoint (line profiles through atoms pairs are shown in Extended Data Figs. 2, 3). From a rigid model structure of the rotated bilayer, assuming that no relaxation occurs (even though some probably does), the model separations for the atom pairs marked in Fig. 3 are predicted to be 0.87 Å, 0.60 Å and 0.36 Å. Although not all atoms can be reconstructed because of scan noise, we have multiple moiré repeats to distinguish random from systematic errors. Ignoring source-size contributions, the expected Rayleigh limit for an incoherent imaging mode for this experimental condition is 1.2 Å, and many atom pairs are completely unresolvable in the ADF image (Extended Data Fig. 3a). Our full-field ptychographic reconstruction demonstrates double the Rayleigh resolution compared to conventional 2α imaging methods. Moreover, some closely spaced atoms lose the central dip at just below 0.40 Å (the Sparrow⁴ criterion for resolution), close to the Abbe limit estimated from Fig. 2.

To understand how dark-field electrons contribute to resolution improvement, we performed additional reconstructions using diffraction patterns with outer cutoff angles varying from one to four times the aperture size $(1\alpha-4\alpha)$. As shown in Fig. 4a–h, when using only the central bright-field disk (1α) , the reconstructed phase (Fig. 4a) has a relatively low resolution, similar to that of the ADF and iCoM images. As the cutoff increases, atoms become sharper and more clearly resolved (Fig. 4a–d). Higher-spatial-frequency information also appears in the diffractograms (Fig. 4e–h). Beyond 3α , where there are fewer scattered electrons, the improvements become less obvious and the reconstruction is limited mainly by the electron dose. As discussed in more detail in the following section, increasing the collection angle beyond the point at which there is meaningful signal in the diffraction pattern does not introduce high-spatial-frequency artefacts. Instead, the reconstruction retains its limiting form. We note that the reconstructed amplitude

of the transmission function also shows the atomic structure of MoS₂. The amplitude modulations are weak, suggesting that the specimen is close to a pure phase object (Extended Data Fig. 4).

As a test of linearity, Fig. 4i shows that the phase at the position of the sulfur monovacancy, where only one sulfur atom is present, is about half of the phase shift of the two-sulfur sites, validating the strong-phase approximation and ePIE reconstruction for these thin 2D materials. That the reconstructed probes (Fig. 4j, k) have similar shapes at different cutoffs also indicates that it is the dark-field electrons that contribute to resolution improvement. The asymmetric probe shape is due to residual aberrations and agrees with measurements using the singular-value-decomposition approach²³.

Influence of electron dose

We explored the potential limits of full-field ptychography further using simulated datasets for a wide range of collection angles and beam currents, including cases where the cutoff is extended beyond most of the scattered electrons and where the dose is too small for a stable reconstruction to be achieved. We evaluate the image quality by both the range of the reconstructed phase and the root-mean-square width of the molybdenum atoms measured from the standard deviation of a Gaussian fit. These measures capture the trends in the height and width of the atom peak, respectively. At high dose, the ultimate information limit of the ptychographic reconstruction is expected to be twice the cutoff angle—that is, 8α for a 4α cutoff. In practice, as shown in Fig. 5, ptychographic reconstructions are influenced mainly by the electron dose, and these limits are not reached for the larger cutoffs. There is only a slight improvement between the 3α and 4α cutoff at a typical operating beam current (1–50 pA), in agreement with the experimental data in Fig. 4. If the beam current is too low (for example, 0.01 pA, which corresponds to a dose of 260 electrons

per Å²), atoms become distorted with reduced image resolution, but some of the overall structure of MoS₂ is still recognizable (Fig. 5c). As the beam current increases, the influence of Poisson noise becomes less important because there are sufficient electrons scattered into high angles to provide interference between higher-order lattice planes. At higher doses, the resolution of the ptychographic reconstruction benefits more fully from the increased maximum collection angle. For large collection angles and high doses, a diminishing return eventually sets in, with the resolution scaling logarithmically with the dose, suggesting that in practice the resolution will ultimately be limited by dose rather than the finite size of the scattering potential (root-mean-square width of about 0.1 Å) or thermal vibrations. Because the EMPAD has a high detective quantum efficiency, increasing the cutoff angle beyond where there is signal does not compromise resolution or introduce additional artefacts, as demonstrated by the fact that all curves in Fig. 5a collapse to the same trend as the beam current decreases.

In Fig. 6 we compare the performance of ePIE and WDD ptychographic reconstructions at low electron dose. Using the same datasets, simulated with a small in-focused probe, both methods yield similar results and can achieve atomic resolution at around 500 electrons per Ų. On the other hand, using a large defocused probe, the ePIE technique can improve reconstruction quality beyond that of WDD (Fig. 6c). Overall, ptychographic reconstructions are more dose-efficient than are low-angle ADF reconstructions (integrating from 1α upwards; Fig. 6d); low-angle ADF is currently the most dose-efficient single-channel STEM imaging mode for single-atom detection 47 . For a more typical range of ADF angles, such as the experimental data of Fig. 2b, the effects are more pronounced (see, for example, Fig. 2d). As shown in Extended Data Fig. 5, the advantage of ptychography over ADF imaging becomes more noticeable for materials with lighter elements, such as graphene.

Discussion

In addition to the beam current and detector configuration required for high-resolution ptychographic reconstruction, other practical sources of errors such as sample contamination and scanning drift may cause distortions and reduce reconstruction quality. However, we have found that full-field ptychography outperforms all other techniques that we have tested under the same conditions (Extended Data Fig. 6). By incorporating other physical constraints and prior knowledge, we envisage that more advanced reconstruction strategies, when applied to full-field electron ptychography data, could compensate for inaccurate scan positions or make allowances for thick specimens with strong dynamic scattering.

In summary, we have demonstrated that with the entire distribution of scattered electrons collected by the EMPAD, full-field ptychography greatly enhances image resolution and contrast compared to traditional electron-imaging techniques, even at low beam voltages. With our improved detector, atomic-scale ptychographic reconstructions are no longer restricted by the aperture size. Instead, image quality is determined by the electron dose and collection angle. Our technique provides an efficient tool for unveiling sub-ångström features of 2D or dose-sensitive materials. Combined with the ultralow-voltage aberration-corrected microscopes that have recently been developed, it has the potential to tackle currently hard problems such as direct imaging of lattice displacements in twisted-layer structures or of structural distortions around single-atom dopants and vacancies, and even three-dimensional tomography.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0298-5.

Received: 14 January 2018; Accepted: 24 May 2018; Published online 18 July 2018.

- Meyer, J. C., Girit, C. O., Crommie, M. F. & Zettl, A. Imaging and dynamics of light atoms and molecules on graphene. Nature 454, 319–322 (2008).
- Krivanek, O. L. et al. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. Nature 464, 571–574 (2010).
- Huang, P. Y. et al. Grains and grain boundaries in single-layer graphene atomic patchwork quilts. Nature 469, 389–392 (2011).
- Sparrow, C. M. On spectroscopic resolving power. Astrophys. J. 44, 76–86 (1916).
- Black, G. & Linfoot, E. H. Spherical aberration and the information content of optical images. *Proc. R. Soc. Lond. A* 239, 522–540 (1957).
- Haider, M. et al. Electron microscopy image enhanced. Nature 392, 768–769 (1998)
- Batson, P. E., Dellby, N. & Krivanek, O. L. Sub-ångström resolution using aberration corrected electron optics. *Nature* 418, 617–620 (2002).
- Erni, R., Rossell, M. D., Kisielowski, C. & Dahmen, U. Atomic-resolution imaging with a sub-50-pm electron probe. *Phys. Rev. Lett.* 102, 096101 (2009)
- Sawada, H. et al. STEM imaging of 47-pm-separated atomic columns by a spherical aberration-corrected electron microscope with a 300-kV cold field emission gun. J. Electron Microsc. 58, 357–361 (2009).
- Kaiser, U. et al. Transmission electron microscopy at 20kV for imaging and spectroscopy. *Ultramicroscopy* 111, 1239–1246 (2011).
- Meyer, J. C. et al. Accurate measurement of electron beam induced displacement cross sections for single-layer graphene. *Phys. Rev. Lett.* 108, 196102 (2012).
- Sawada, H., Sasaki, T., Hosokawa, F. & Suenaga, K. Atomic-resolution STEM imaging of graphene at low voltage of 30 kV with resolution enhancement by using large convergence angle. *Phys. Rev. Lett.* 114, 166102 (2015).
- Linck, M. et al. Chromatic aberration correction for atomic resolution TEM imaging from 20 to 80 kV. Phys. Rev. Lett. 117, 076101 (2016).
- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q. Rev. Biophys. 28, 171–193 (1995).
- 15. Gabor, D. A new microscopic principle. Nature 161, 777-778 (1948).
- Hoppe, W. Beugung im inhomogenen Primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeungungsinterferenzen. Acta Crystallogr. A 25, 495–501 (1969).
- Nellist, P. D., McCallum, B. C. & Rodenburg, J. M. Resolution beyond the 'information limit' in transmission electron microscopy. *Nature* 374, 630–632 (1995).
- Nellist, P. D. & Rodenburg, J. M. Electron ptychography. I. Experimental demonstration beyond the conventional resolution limits. Acta Crystallogr. A 54, 49–60 (1998).
- Thibault, P. et al. High-resolution scanning X-ray diffraction microscopy. Science 321, 379–382 (2008).
- Maiden, A. M. & Rodenburg, J. M. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* 109, 1256–1262 (2009).
- Thibault, P. & Menzel, A. Reconstructing state mixtures from diffraction measurements. *Nature* 494, 68–71 (2013).
- Li, P., Edo, T. B. & Rodenburg, J. M. Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design. *Ultramicroscopy* 147, 106–113 (2014).
- Yang, H. et al. Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures. Nat. Commun. 7, 12532 (2016).
- Pelz, P. M., Qiu, W. X., Bücker, R., Kassier, G. & Miller, R. J. D. Low-dose cryo electron ptychography via non-convex Bayesian optimization. Sci. Rep. 7, 9883 (2017)
- Maiden, A. M., Humphry, M. J. & Rodenburg, J. M. Ptychographic transmission microscopy in three dimensions using a multi-slice approach. J. Opt. Soc. Am. A 29, 1606–1614 (2012).
- Gao, S. et al. Electron ptychographic microscopy for three-dimensional imaging. Nat. Commun. 8, 163 (2017).
- Rodenburg, J. M., Hurst, A. C. & Cullis, A. G. Transmission microscopy without lenses for objects of unlimited size. *Ultramicroscopy* 107, 227–231 (2007).
- Rodenburg, J. M. et al. Hard-X-ray lensless imaging of extended objects. *Phys. Rev. Lett.* 98, 034801 (2007).
- Hüe, F., Rodenburg, J. M., Maiden, A. M., Sweeney, F. & Midgley, P. A. Wave-front phase retrieval in transmission electron microscopy via ptychography. *Phys. Rev. B* 82, 121415 (2010).
- Hüe, F., Rodenburg, J. M., Maiden, A. M. & Midgley, P. A. Extended ptychography in the transmission electron microscope: possibilities and limitations. *Ultramicroscopy* 111, 1117–1123 (2011).
- Putkunz, C. T. et al. Atom-scale ptychographic electron diffractive imaging of boron nitride cones. *Phys. Rev. Lett.* 108, 073901 (2012).
- D'Alfonso, A. J. et al. Deterministic electron ptychography at atomic resolution. Phys. Rev. B 89, 064101 (2014).
- Pennycook, T. J. et al. Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: experimental demonstration at atomic resolution. Ultramicroscopy 151, 160–167 (2015).
- Yang, H. et al. Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution. *Ultramicroscopy* 180, 173–179 (2017).
- Wang, P., Zhang, F., Gao, S., Zhang, M. & Kirkland, A. I. Electron ptychographic diffractive imaging of boron atoms in LaB₆ crystals. Sci. Rep. 7, 2857 (2017).



- Humphry, M. J., Kraus, B., Hurst, A. C., Maiden, A. M. & Rodenburg, J. M. Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging. *Nat. Commun.* 3, 730 (2012).
- Frojdh, E. et al. Count rate linearity and spectral response of the Medipix3RX chip coupled to a 300μm silicon sensor under high flux conditions. J. Instrum. 9, C04028 (2014).
- 38. Rose, A. Vision Human and Electronic Ch. 1 (Plenum Press, New York, 1949).
- Tate, M. W. et al. High dynamic range pixel array detector for scanning transmission electron microscopy. *Microsc. Microanal.* 22, 237–249 (2016).
- Close, R., Chen, Z., Shibata, N. & Findlay, S. D. Towards quantitative, atomicresolution reconstruction of the electrostatic potential via differential phase contrast using electrons. *Ultramicroscopy* 159, 124–137 (2015).
- Lazić, I., Bosch, E. G. T. & Lazar, S. Phase contrast STEM for thin samples: integrated differential phase contrast. *Ultramicroscopy* 160, 265–280 (2016).
- Maiden, A. M., Humphry, M. J., Zhang, F. & Rodenburg, J. M. Superresolution imaging via ptychography. J. Opt. Soc. Am. A 28, 604–612 (2011).
- Rodenburg, J. M. & Bates, R. H. T. The theory of super-resolution electron microscopy via Wigner-distribution deconvolution. *Philos. Trans. R. Soc. Lond. A* 339, 521–553 (1992).
- 44. Lee, J. & Barbastathis, G. Denoised Wigner distribution deconvolution via low-rank matrix completion. *Opt. Express* **24**, 20069–20079 (2016).
- 45. Abbe, E. The relation of aperture and power in the microscope. *J. R. Microsc. Soc.* **2**, 300–309 (1882).
- van der Zande, A. M. et al. Tailoring the electronic structure in bilayer molybdenum disulfide via interlayer twist. *Nano Lett.* 14, 3869–3875 (2014).
- Hovden, R. & Muller, D. A. Efficient elastic imaging of single atoms on ultrathin supports in a scanning transmission electron microscope. *Ultramicroscopy* 123, 59–65 (2012).

Acknowledgements Y.J. and V.E. acknowledge support from DOE grant DE-SC0005827. Z.C. and D.A.M. are supported by the PARADIM Materials Innovation Platform in-house programme by NSF grant DMR-1539918.

We acknowledge the electron microscopy facility and support for Y.H. from the NSF MRSEC programme (DMR 1719875) and NSF MRI grant DMR-1429155. H.G., S.X. and J.P. acknowledge additional support from AFOSR MURI (FA9550-16-1-003) and UChicago NSF MRSEC programme (DMR 1420709). Detector development at Cornell was supported by the Kavli Institute at Cornell for Nanoscale Science and DOE grant DE-SC0017631 to S.M.G. We thank K. Nguyen, P. Huang, M. Humphry and P. Nellist for discussions and B. Jiang from Thermo Scientific for help during the initial experiments.

Reviewer information *Nature* thanks J. Rodenburg and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Experiments were performed and designed by Z.C., Y.H. and D.A.M. Y.J. contributed to data analysis and ptychographic reconstruction, with support from V.E. Sample preparation was done by P.D. and Y.H., from MoS₂ thin films synthesized by H.G., S.X. and J.P. EMPAD was optimized by P.P., M.W.T. and S.M.G. All authors discussed the results and implications throughout the investigation. All authors have approved the final version of the manuscript.

Competing interests Cornell University has licensed the EMPAD hardware to Thermo Scientific.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0298-5.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0298-5

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.A.M. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



METHODS

EMPAD data acquisition. The 4D dataset of monolayer MoS₂ was taken using an aberration-corrected FEI Titan with 8.2-pA beam current, 80-keV beam energy and 21.4-mrad aperture size, with the dose limited by the radiation resistance of the sample. The EMPAD has 128×128 pixels and a readout speed of 0.86 ms per frame. The exposure time was set to 1 ms per frame in all experiments. 51×87 diffraction patterns with a scan step size of 0.21 Å were used to generate brightfield, ADF and iCoM images in Fig. 2a–c. The ADF image was integrated from 64.2 mrad (3 α) to 84.6 mrad (4 α). Higher angles did not add important contributions to the signal. The dataset of twisted MoS₂ in Fig. 3 was taken with the same beam conditions except for a 10.1-pA beam current. 68 \times 68 diffraction patterns with a scan step size of 0.59 Å were used for ptychographic reconstruction.

Ptychographic reconstructions. Before reconstruction, all diffraction patterns are padded with zeros to a total size of 256×256 and thus the pixel size in the reconstructed phase is 0.12 Å per pixel. The ePIE method²⁰ is implemented with modifications to exclude bad pixels in the diffraction patterns. The algorithm uses a multiplicative approximation, which is a generalization of the strong-phase approximation, allowing for both a strong phase object and a variable amplitude term. It aims to minimize the Euclidian distance between reconstructed and measured diffraction patterns. In general, the convergence of reconstructions depends on the number of iterations and update parameters for the transmission function and the probe function. Because experimental data contain noise and other sources of errors, fast convergence may introduce noisy artefacts and reduce reconstruction quality^{48,49}. To alleviate this problem, we used a small update parameter (0.1) for the transmission function and limited the reconstruction of the probe function to data taken in areas with minimal contamination. For our experimental conditions, with cutoff angle θ_{max} = 80 mrad, the thickness limit dz for treating the sample as a projected potential and neglecting beam propagation is estimated to be $dz = \lambda/[2\sin^2(\theta_{max}/2)] \approx 1.3 \text{ nm}^{50}$, which is larger than the thicknesses of both the monolayer MoS_2 (3.1 Å) and the bilayer MoS_2 (9.8 Å).

Fourier resolution estimate. All diffractograms (Fourier intensity) of brightfield, ADF, iCoM and ptychography reconstructions in Fig. 2 were calculated from images constructed from 128×128 diffraction patterns (see Extended Data Fig. 6). To visualize diffraction spots, a periodic and smooth decomposition⁵¹ was applied to images to reduce artefacts caused by edge discontinuities. Next, the real-space image was multiplied by a Gaussian function, making the diffraction spots slightly larger and thus more visible. The Fourier intensity was rescaled to enhance the intensity of higher-order spots for better visualization.

4D data simulation. All datasets used for dose-cutoff simulations were generated by the μ STEM software⁵², which models the atomic potential using scattering

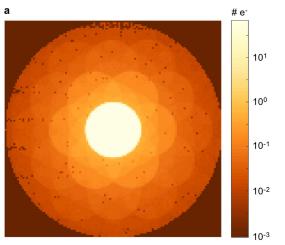
factors given in ref. 53 . For the in-focused probe, 21×24 diffraction patterns with a 0.45-Å scan step size were simulated at 80-keV beam energy and with 21.4-mrad aperture size. The thermal diffuse scattering effect was included with the frozen-phonon approximation. The diffraction patterns were further corrupted with Poisson noise determined by the simulated beam dose. Extended Data Fig. 7 shows selected ePIE reconstructions at different beam currents (0.01–100 pA) and cutoff angles $(1\alpha-4\alpha)$. Simulations (Fig. 6c) of the large probe profile used a 80-nm defocus and 10.04-Å scan step size.

Effect of chromatic aberrations. Ptychography is less sensitive to chromatic blur than is direct phase imaging 54 , but it is still sensitive, especially at low doses. 4D datasets including chromatic aberrations were simulated for a Gaussian energy spread with full-width at half-maximum of $\Delta E=1.1$ eV, a chromatic aberration coefficient of $C_{\rm c}=1.72$ mm and a beam energy of 80 keV. Two different probe-forming aperture sizes, 21.4 mrad and 35 mrad, were chosen to reflect conditions under which chromatic blur is moderate and strong, respectively. The dose dependence of ePIE reconstructions for datasets with and without chromatic aberration is shown in Extended Data Fig. 8. Under the experimental conditions, the influence of chromatic aberration is visually negligible. At lower doses, the larger chromatic blur for the larger aperture leads to a worse reconstruction.

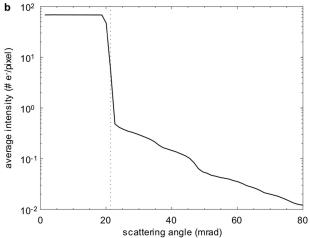
Data availability. All relevant data are available from the corresponding author on request.

Code availability. Code developed at Cornell, including visualization software for 4D datasets, is available from the corresponding author on request. $\mu STEM$ was developed at the University of Melbourne and can be downloaded from http://tcmp.ph.unimelb.edu.au/mustem/muSTEM.html.

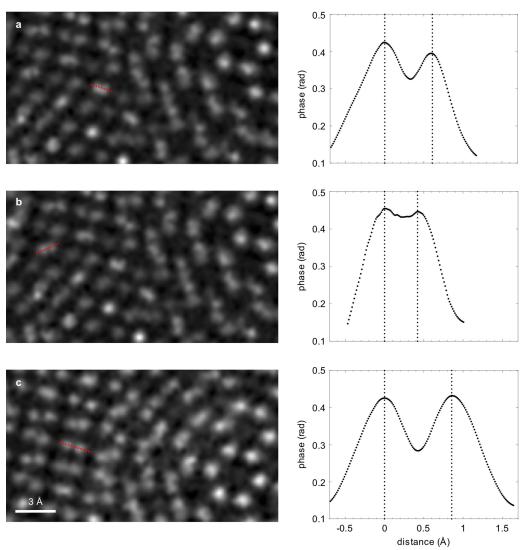
- Zuo, C., Sun, J. & Chen, Q. Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy. Opt. Express 24, 20724–20744 (2016).
- 49. Maiden, A., Johnson, D. & Li, P. Further improvements to the ptychographical iterative engine. *Optica* **4**, 736–745 (2017).
- Suzuki, A. et al. High-resolution multislice X-ray ptychography of extended thick objects. *Phys. Rev. Lett.* 112, 053903 (2014).
- Hovden, R., Jiang, Y., Xin, H. L. & Kourkoutis, L. F. Periodic artifact reduction in Fourier transforms of full field atomic resolution images. *Microsc. Microanal.* 21, 436–441 (2015).
- Allen, L. J., D'Alfonso, A. J. & Findlay, S. D. Modelling the inelastic scattering of fast electrons. *Ultramicroscopy* 151, 11–22 (2015).
- Waasmaier, D. & Kirfel, A. New analytical scattering-factor functions for free atoms and ions. Acta Crystallogr. A 51, 416–431 (1995).
- Nellist, P. D. & Rodenburg, J. M. Beyond the conventional information limit: the relevant coherence function. *Ultramicroscopy* 54, 61–74 (1994).



Extended Data Fig. 1 | Position-averaged diffraction pattern of the 4D dataset from monolayer MoS₂. a, Position-averaged convergent beam electron diffraction (CBED) pattern from the 4D dataset from monolayer



MoS₂. **b**, Radially averaged intensity distribution (on a logarithmic scale) of the CBED pattern, showing the dynamic range spanned by the scattering distribution.

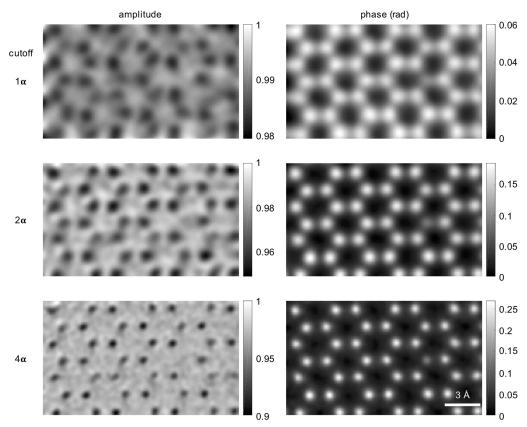


Extended Data Fig. 2 | Line profiles through atom pairs in the twisted bilayer MoS_2 . Line profiles are from atom pairs in Fig. 3, with the respective subregions shown on the left. a-c, The measured peak–peak

separations between two atoms are 0.42 \pm 0.02 Å (a), 0.61 \pm 0.02 Å (b) and 0.85 \pm 0.02 Å (c).

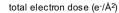
Extended Data Fig. 3 | ADF image and line profiles through atom pairs in the twisted bilayer MoS₂. a, ADF image synthesized from the 4D diffraction dataset. b, Phase of the transmission function reconstructed by ptychography. The yellow marker indicates a pair of atoms that is predicted to have a separation of 0.2 Å on the basis of the structural model,

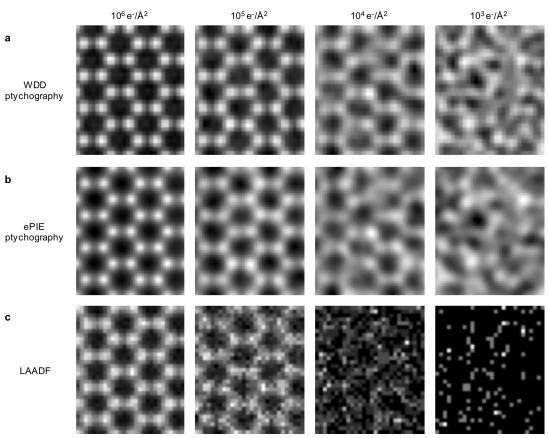
but cannot be resolved explicitly in our reconstruction. For a more detailed comparison, a red box is placed over corresponding regions in **a** and **b**. **c**, Enlarged image of the red boxed region in **b**, with the false colour scale of Fig. 3. **d**, Line profiles across the atom pairs labelled with dashed lines in **c**. The peak–peak separations are overlaid near the line profiles.



Extended Data Fig. 4 | Reconstructed amplitude and phase of monolayer MoS_2 at different cutoff angles. Both the amplitude (left panels) and phase (right panels) of the reconstructed transmission function show the atomic structure of monolayer MoS_2 . Image resolution

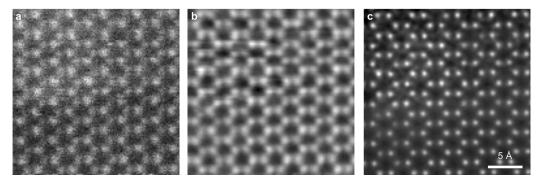
improves as the cutoff angle increases. Amplitude modulations are relatively weak, deviating by only a few per cent from a pure phase object (that is, an object function with unit amplitude).





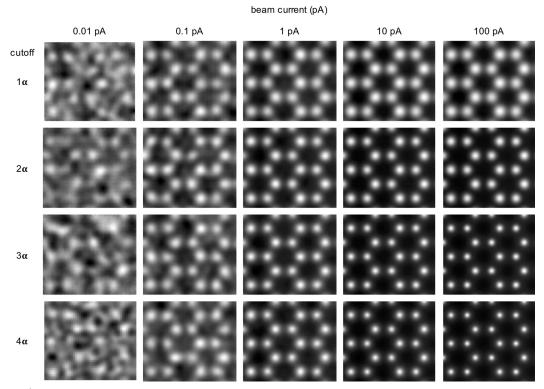
Extended Data Fig. 5 | Comparison between ptychography techniques and low-angle ADF imaging of graphene. a, b, Ptychographic reconstructions of simulated data with an in-focused probe, using the WDD (a) and ePIE (b) methods. c, Low-angle ADF (integrating from

 1α to $4\alpha)$ reconstruction using the same simulated datasets. Both ptychographic methods show similar reconstructions and are about 10 times more dose-efficient than the low-angle ADF technique. Beam energy, 80 keV; aperture size (α) , 21.4 mrad.



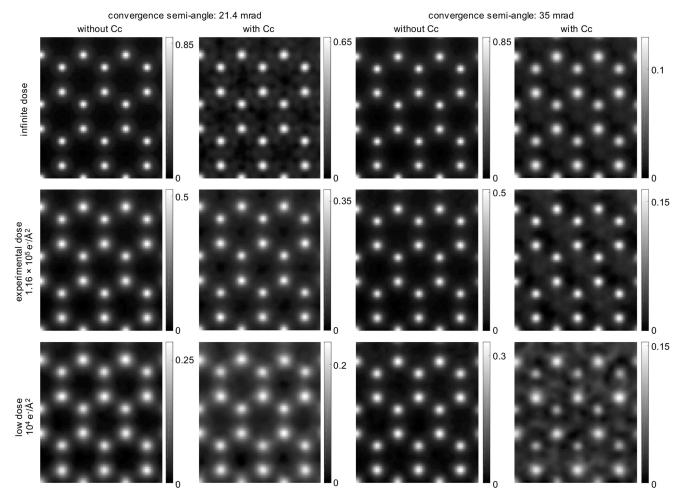
Extended Data Fig. 6 | Influence of scanning drift and contamination. a–c, ADF image (a), iCoM image (b) and phase of transmission reconstructed by full-field ptychography (c) using 128×128 diffraction patterns, covering a field of view of 2.7 nm $\times 2.7$ nm. The ADF and iCoM reconstructions both suffer from stripe artefacts and large contrast

variations. In the ptychographic reconstruction, scanning drift distorts and blurs reconstructed atoms in the vicinity of the scan distortion, but the overall resolution away from the distortion remains higher than the other imaging modes.



Extended Data Fig. 7 | Effect of dose and cutoff angles on ptychographic reconstructions of monolayer MoS_2 using simulated diffraction patterns. At high beam current, the resolution of the ptychography reconstruction is fundamentally determined by the collection angle of the

detector. As the beam current decreases, the resolution becomes dose-limited and noise artefacts start to appear in the ePIE reconstruction. Beam energy, 80 keV; aperture size (α), 21.4 mrad.



Extended Data Fig. 8 | Effect of chromatic aberrations at different electron doses for ptychographic reconstructions of monolayer MoS₂ using simulated datasets at 80 keV. Two convergence semi-angles are shown, 21.4 mrad (left two columns) and 35 mrad (right two columns), representing conditions under which chromatic aberrations have moderate and large effects on the incident probe shape, respectively (C_c = 1.72 mm, ΔE = 1.1 eV). 21.4 mrad is also the experimental convergence angle. The incident electron dose levels are an infinite dose (top row), the experimental dose of 1.16 \times 10⁵ electrons per Å² (middle row) and a

low dose of 10^4 electrons per Å 2 (bottom row). In the presence of noise, chromatic aberrations degrade the phase range of the reconstruction compared with the achromatic data. The data for the larger convergence semi-angle are more strongly affected. At infinite and experimental doses, ptychographic reconstructions with and without chromatic aberration are visually similar for both convergence angles. At low dose and with chromatic aberration, the reconstructed atoms are broadened, and distinct artefacts appear for a convergence angle of 35 mrad.



Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations

Po-Ru Loh 1,2,14* , Giulio Genovese 2,3,4,14* , Robert E. Handsaker 2,3,4 , Hilary K. Finucane 2,5 , Yakir A. Reshef 6 , Pier Francesco Palamara 7 , Brenda M. Birmann 8 , Michael E. Talkowski 2,3,9,10 , Samuel F. Bakhoum 11,12 , Steven A. McCarroll 2,3,4,15* & Alkes L. Price 2,13,15*

The selective pressures that shape clonal evolution in healthy individuals are largely unknown. Here we investigate 8,342 mosaic chromosomal alterations, from 50 kb to 249 Mb long, that we uncovered in blood-derived DNA from 151,202 UK Biobank participants using phase-based computational techniques (estimated false discovery rate, 6-9%). We found six loci at which inherited variants associated strongly with the acquisition of deletions or loss of heterozygosity in cis. At three such loci (MPL, TM2D3-TARSL2, and FRA10B), we identified a likely causal variant that acted with high penetrance (5-50%). Inherited alleles at one locus appeared to affect the probability of somatic mutation, and at three other loci to be objects of positive or negative clonal selection. Several specific mosaic chromosomal alterations were strongly associated with future haematological malignancies. Our results reveal a multitude of paths towards clonal expansions with a wide range of effects on human health.

Clonal expansions of blood cells containing somatic mutations are often observed in individuals without cancer $^{1-13}$. Consistent with the idea that clonal mosaicism can be a precancerous state, detectable mosaicism confers a more than tenfold increased risk of future haematological malignancy $^{1-4}$ and often involves pro-proliferative mutations. Several studies have suggested that inherited variation can influence the likelihood of clonal mosaicism $^{11,14-21}$.

The limiting factor in almost all studies of clonal mosaicism has been sample size, with earlier insights arising from analyses of up to around 1,000 mosaic events. Two key factors determine the number of detectable mosaic mutations: the number of individuals analysed, and the ability to detect clonal expansions present at low-to-modest cell fractions. Here we describe insights from an analysis of 8,342 mosaic chromosomal alterations (mCAs) which we identified in single nucleotide polymorphism (SNP) array data from 151,202 UK Biobank participants²² using a sensitive algorithm we developed to make use of long-range haplotype phase information (building on published work⁸). We also draw upon data on health outcomes during 4–9 years after DNA sampling.

These data provide insights into clonal expansion, including mechanisms by which inherited variants at several loci act in *cis* to generate or propel mosaicism. We also identify specific mCAs that associate strongly with future haematological malignancies.

Mosaic chromosomal alterations in UK Biobank

We analysed allele-specific SNP-array intensity data previously obtained by genotyping blood-derived DNA from 151,202 UK Biobank participants (40–70 years of age)²²; 607,525 genotyped variants remained after quality control (see Methods). We detected mCAs at cell fractions as low as 1% by using long-range phase information

that is uniquely available in the UK Biobank^{23,24}. Intuitively, accurate phasing allows the detection of subtle imbalances in the abundances of two haplotypes by combining allele-specific information across a very large number of SNPs (Extended Data Fig. 1). To make maximal use of phase information, we developed a new statistical method for phase-based mCA detection (see Methods and Supplementary Note 1).

We detected 8,342 mCAs (in 7,484 of the 151,202 individuals analysed) at an estimated false discovery rate (FDR) of 6-9% (Fig. 1, Extended Data Fig. 2, Supplementary Table 1, and Supplementary Notes 2, 3; validation rates could differ from this FDR estimate). We confidently classified 71% of the detected mCAs as either loss, copy-number neutral loss of heterozygosity (CNN-LOH), or gain; for the other 29% of events, copy-number state could not be inferred definitively (Fig. 2a and Supplementary Note 1). Most detected mCAs (5,901 of 8,342) were present at inferred cell fractions below 5% (Supplementary Note 4) and would have been undetectable without long-range phasing (Supplementary Note 5). The genomic distribution of detected mCAs was broadly consistent with those found in previous studies^{1,2,7,8}, as was the observation that individuals acquire multiple mCAs much more frequently than expected by chance (Fig. 2b, Extended Data Fig. 3, Supplementary Tables 2, 3, and Supplementary Note 6); differences (for example, in relative rates of del(20q) calls²⁵) could be explained by differing methodological sensitivity or genotyping platforms (Supplementary Note 4).

Commonly deleted regions (CDRs) below 1 Mb in length may indicate haploinsufficient tumour-suppressor genes for which loss of one copy promotes cell proliferation². Focal deletions most frequently targeted 13q14, *DNMT3A*, and *TET2*, as previously observed^{2,8}; we further observed that most CNN-LOH events on 13q, 2p, and 4q spanned the same CDRs (Fig. 1 and Supplementary Note 7). We

¹Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ²Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Department of Genetics, Harvard Medical School, Boston, MA, USA. ⁵Schmidt Fellows Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁵Department of Computer Science, Harvard University, Cambridge, MA, USA. ⁷Department of Statistics, University of Oxford, Oxford, UK. ⁸Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ⁹Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ¹⁰Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ¹¹Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹²Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. ¹³Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ¹⁴These authors contributed equally: Po-Ru Loh, Giulio Genovese. ¹⁵These authors jointly supervised this work: Steven A McCarroll, Alkes L Price. *e-mail: poruloh@broadinstitute.org; giulio.genovese@gmail.com; mccarroll@genetics.med.harvard.edu

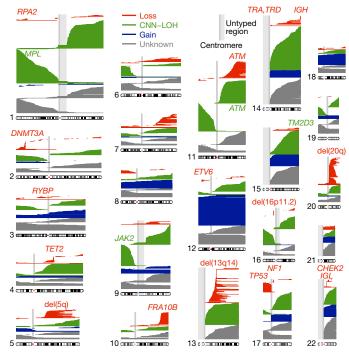


Fig. 1 | Mosaic chromosomal alterations detected in 151,202 UK Biobank participants. Each horizontal line corresponds to an mCA; a total of 5,562 autosomal events in 4,889 unique individuals are displayed. We detected an additional 2,780 chromosome X events in females (mostly whole-chromosome losses). Detected events are colour-coded by copy number. Focal deletions are labelled in red with names of putative target genes. Loci containing inherited variants influencing somatic events in *cis* are labelled in the colour of the mCA (red for del(10q)-associated *FRA10B*, green for CNN-LOH-associated loci). Enlarged per-chromosome plots are provided in Supplementary Note 2.

detected new CDRs at *ETV6*, *NF1*, and *CHEK2*, which are commonly mutated in cancers, and at *RPA2* and *RYBP*. We also detected a CDR at 16p11.2 overlapping a region whose deletion is a known risk factor for autism and other neuropsychiatric phenotypes, though we did not detect this mCA among 2,079 sequenced genomes from the Simons Simplex Collection (SSC)^{26,27} (Supplementary Note 8). Deletions tended to be concentrated on chromosomes that are seldom duplicated²⁸ (Fig. 2c and Supplementary Table 1), supporting the theory that cumulative haploinsufficiency and triplosensitivity shape clonal evolution²⁹.

We found several notable exceptions to a general pattern in which acquired mutations are most common in the elderly and in males^{1,2,7,8} (Fig. 2d and Supplementary Table 4). Loss of chromosome X in females³⁰ was by far the most common event we detected (Supplementary Table 1 and Supplementary Note 2), with frequency increasing markedly with advancing age (Fig. 2d and Supplementary Table 4). (We did not examine loss of chromosome Y, which has been studied elsewhere²¹.) Stratification of autosomal mCAs by location and copy number revealed an unexpected relationship: although most gain events were (as expected) enriched in elderly individuals and in males, CNN-LOH events tended to affect both sexes equally (Fig. 2e and Supplementary Table 5). Three mCAs exhibited unusual age and sex distributions (FDR 0.05; binomial and z-tests): gains on chromosome 15 were much more frequent in elderly males³¹; and 16p11.2 deletions and 10q terminal deletions were much more frequent in females and exhibited frequency unrelated to age. Age-independent events could in principle occur early in development or take less time to reach high cell fractions; sex-specific effects (which we replicated in previous data sets^{1,2,8}; Supplementary Note 3) will require future work to explain.

Some acquired mutations could in principle arise or be selected within specific haematopoietic lineages. We tested this hypothesis

by examining individuals in the top percentile for counts of lymphocytes, basophils, monocytes, neutrophils, red blood cells, or platelets. We identified many mCAs that were significantly concentrated (FDR 0.05; Fisher's exact test) in one or more of these subsets of the cohort (Fig. 2f and Supplementary Table 6). Consistent with the idea that these relationships might reflect clonal selection in specific blood cell types, mutations commonly observed in chronic lymphocytic leukaemia (CLL) 32,33 were enriched among individuals with high lymphocyte counts, and *JAK2*-related 9p events (which are commonly observed in myeloproliferative neoplasms (MPNs)) were most common among individuals with high myeloid cell counts. While future work will be needed to replicate and further explore these findings, our results suggest that mCAs may produce blood-composition phenotypes in individuals with no known malignancy.

Inherited variants affect acquisition of nearby mCAs

To identify inherited influences on mCA formation or selection, we performed chromosome-wide scans for associations between mCAs and germline variants on the same chromosome (see Methods). This analysis revealed four loci at which inherited variation strongly associated with the acquisition of genomically nearby autosomal mCAs, and two loci on chromosome X associated with X loss in females (Table 1, Figs. 3, 4). We also replicated an earlier association of the *JAK2* 46/1 haplotype with 9p CNN-LOH^{15-18,20} (Extended Data Fig. 4). To identify mechanisms that might underlie these associations, we finemapped these loci using whole-genome sequence (WGS) data and studied the phase of risk alleles relative to associated chromosomal alterations in *cis*.

Somatic terminal 10q deletions associated strongly ($P = 6.1 \times 10^{-42}$; Fisher's exact test) with the common SNP rs118137427 near FRA10B, a known genomic fragile site^{34,35} at the estimated common breakpoint of the 10q deletions (Table 1 and Fig. 3a). All 60 individuals with these mosaic 10q deletions had inherited the rs118137427:G risk allele (the allele frequency is 5% in the population), which was always inherited on the same chromosome that subsequently acquired a terminal deletion (Table 1).

To identify a causal variant potentially tagged by the rs118137427:G risk allele, we searched for acquired 10q deletions in WGS data from 520 SSC families (see Methods). We identified two parent-child duos in which both parent and child had acquired the 10q terminal deletion (in mosaic form); all four individuals possessed expanded AT-rich repeats at FRA10B on the rs118137427:G haplotype background (P = 0.01; Fig. 3c). Further evidence that the rs118137427:G risk allele tags an unstable version of the FRA 10B locus³⁶ was provided by analysis of the variable number tandem repeat (VNTR) sequence at FRA 10B (in all 2,079 individuals). This analysis revealed a diversity of novel VNTR sequence motifs (12 distinct primary repeat units carried by 26 individuals from 14 families), all on the rs118137427:G haplotype background (Extended Data Fig. 5a, b and Supplementary Note 8). (The VNTR motifs did not associate with autism status in the SSC cohort.) The motifs had lengths of 38, 39, 42, and 43 bp and exhibited evidence of repeat expansion (probably more than 75 copies in the longest alleles³⁵); by contrast, the hg19 reference sequence at FRA10B contains three copies of a 40-bp repeat. Imputing the VNTRs into the UK Biobank showed that they explained 24 of 60 del(10q) cases, despite being present in only about 0.7% of the cohort (Supplementary Table 7). Notably, individuals with del(10q) were as young as other UK Biobank participants, and 51 of 60 were female (binomial $P = 1.8 \times 10^{-7}$) (Fig. 3b); these unusual patterns (which were shared with 16p11.2 deletions) will require further study (Supplementary Note 8).

CNN-LOH events on chromosome (chr)1p strongly associated ($P=6.2\times10^{-16}$, lead SNP rs144279563) with three independent, rare risk haplotypes (allele frequencies = 0.01–0.05%) at the MPL proto-oncogene at 1p34.1; the three haplotypes increased risk for 1p CNN-LOH by factors of 53, 63, and 103 (95% confidence intervals (CIs): 28–99, 29–139, and 35–300, respectively) (Table 1, Fig. 4a, and Supplementary Table 8). Other individuals with 1p CNN-LOH mosaicism also shared

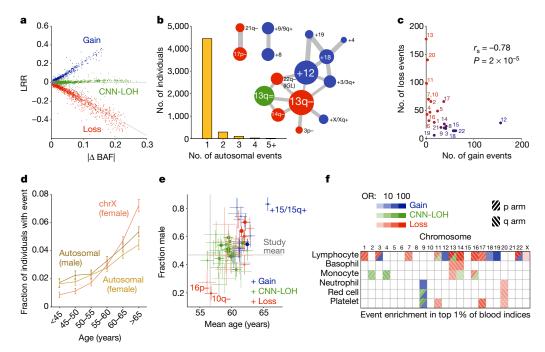


Fig. 2 | Distributional properties of detected mCAs. a, $\log_2 R$ ratio (LRR), measuring total allelic intensity, scales roughly linearly with B-allele frequency (BAF) deviation, measuring relative allelic intensity, among events with each copy number^{1,2,8}. b, Most individuals with a detected autosomal mCA have only one event, although a larger number than expected (441 versus 100) have multiple events. Several pairs of mCA types co-occur much more frequently than expected by chance; edge weights in the co-occurrence graph scale with enrichment. c, Autosomes with more gain events tend to have fewer loss events (excluding deletions involving V(D)J recombination on chromosomes 14 and 22); Spearman's

test on n=22 autosomes. **d**, Fractions of individuals with at least one detected autosomal event increase steadily with age, and this trend is even more pronounced for X chromosome events in females. Error bars, 95% CI. **e**, Carriers of different mCA types have different age and sex distributions. Error bars, s.e.m. **f**, Different mCAs are significantly enriched (FDR 0.05) among individuals with anomalous blood counts in different blood lineages (adjusted for age, sex, and smoking status; see Methods). Numeric data including exact sample sizes used to compute error bars are provided in Supplementary Tables 1–6.

long haplotypes containing *MPL*, suggesting the existence of additional very rare risk variants (Extended Data Fig. 5c). Notably, although gain-of-function mutations in *MPL* lead to myeloproliferative neoplasms^{37,38}, the lead SNP on one haplotype, rs369156948, is a protein-truncating variant (PTV) in *MPL* with no association to haematological malignancies in the UK Biobank (0 cases among 36 carriers).

We were able to identify a likely mechanism for selection of the CNN-LOH events involving *MPL*. For all 16 events for which we could

confidently phase the inherited risk allele relative to the somatic CNNLOH, the CNN-LOH mutation had replaced the clonal haematopoiesis risk allele with the reference allele (binomial $P=3\times 10^{-5}$; Table 1 and Fig. 4a). These results suggest that, among individuals with rare inherited variants that reduce MPL function, recovery of normal MPL gene activity via CNN-LOH provides a proliferative advantage.

CNN-LOH events on chr11q associated ($P = 7.4 \times 10^{-9}$, OR = 41 (18–94)) with a rare risk haplotype (allele frequency = 0.07%)

Table 1 | Novel genome-wide significant associations of mCAs with inherited variants

SV type	Locus	Variant	Location	Allelesa	RAF ^b	GWAS		Risk allelic shift in hets		
						Р	OR (95% CI)	N _{inc} c	$N_{\rm dec}{}^{\rm d}$	P
cis associations										
10q loss	FRA10B	rs118137427e	10q25.2	A/G	0.05	6.1×10^{-42}	18 (12-26)	0	43	2.3×10^{-13}
1p CNN-LOH	MPL	rs144279563	1p34.1	C/T	0.0005	6.2×10^{-16}	53 (28–99)	0	9	3.9×10^{-3}
		rs182971382	1p34.1	A/G	0.0003	3.0×10^{-11}	63 (29–139)	0	4	1.3×10^{-1}
		rs369156948 ^f	1p34.2	C/T	0.0001	7.3×10^{-8}	103 (35-300)	0	3	2.5×10^{-1}
11q CNN-LOH	ATM	rs532198118	11q22.3	A/G	0.0007	7.4×10^{-9}	41 (18–94)	6	0	3.1×10^{-2}
15q CNN-LOH and loss	TM2D3, TARSL2	70 kb deletion ^g	15q26.3	CN = 1/0	0.0003	1.3×10^{-86}	698 (442–1102)	39	2	7.8×10^{-10}
chrX loss	DXZ1	rs2942875	Xp11.1	T/C	0.55	9.7×10^{-4}	1.09 (1.04–1.15)	423	796	6.6×10^{-27}
	DXZ4	rs11091036	Xq23	C/G	0.73	1.1×10^{-3}	1.10 (1.04–1.17)	369	555	1.0×10^{-9}
trans associations										
chrX loss	SP140L	rs725201	2q37.1	G/T	0.56	9.2×10^{-10}	1.17 (1.12-1.24)			
	HLA	rs141806003	6p21.33	C/CAAAG	0.34	6.1×10^{-10}	1.18 (1.12–1.25)			

Results of two independent association tests are reported: a Fisher test treating individuals with a given mCA type as cases; and (for cis associations) a binomial test for biased allelic imbalance in heterozygous cases (hets; see Methods). All loci reaching $P < 1 \times 10^{-8}$ in either test are reported; each cis association detected by one test reached nominal (P < 0.05) significance in the other test. At significant loci, the lead associated variant as well as additional independent associations reaching $P < 1 \times 10^{-6}$ are reported.

*Risk-lowering/risk-increasing allele.

^bRisk allele frequency (in UK Biobank participants with European ancestry).

Number of mosaic individuals heterozygous for the variant in which the somatic event shifted the allelic balance in favour of the risk allele (by duplication of its chromosomal segment and/or loss of the homologous segment).

^dNumber of mosaic individuals heterozygous for the variant in which the somatic event shifted the allelic balance in favour of the non-risk allele.

ers118137427 tags expanded repeats at FRA10B (Fig. 3).

frs369156948 is a nonsense mutation in MPL.

gThis deletion spans chr15:102.15–102.22Mb (hg19) and is tagged by rs182643535.

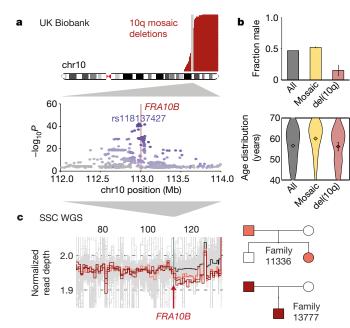


Fig. 3 | **Repeat expansions at fragile site** *FRA10B* **driving breakage at 10q25.2. a**, Germline variants at 10q25.2 associate strongly with terminal 10q mosaic deletion (Fisher's exact test, n=120,664 individuals). Left boundaries of the deletions are called with error; true breakpoints are probably near-identical (Supplementary Note 4). b, UK Biobank carriers of terminal 10q deletion are predominantly female (top; 51 of n=60 individuals; error bars, 95% CI) with age distribution similar to the overall study population (bottom; violin plot centres, means; error bars, 95% CI). **c**, WGS samples with terminal 10q deletion (two parent–child duos; right) carry inherited expanded repeats at *FRA10B*.

surrounding the *ATM* gene at 11q22.3 (Table 1, Fig. 4b, and Supplementary Table 8). For all six CNN-LOH events for which we could confidently phase the risk allele relative to the somatic mutation, the LOH mutation had caused the rare risk allele to become homozygous, suggesting that the risk allele confers a proliferative advantage in the homozygous state (Table 1 and Fig. 4b). (This dynamic contrasts with that of *MPL*, at which the rare, inherited risk haplotypes were eliminated by LOH and clonal selection.) While sequencing would be required to identify a causal variant, *ATM* is a clear putative target: *ATM* encodes a DNA-damage response kinase that promotes DNA repair and limits cell division, and *ATM* is often inactivated by mutation in cancers^{32,33}. In our analysis, acquired 11q deletions also appeared to target *ATM* (Fig. 1 and Supplementary Note 2).

CNN-LOH and loss events at chr15q associated strongly $(P = 1.3 \times 10^{-86})$ with a rare, inherited 70-kb deletion (allele frequency = 0.03%) that spanned all of TM2D3 and part of TARSL2 at 15q26.3 (Table 1, Fig. 4c, and Extended Data Figs. 6, 7). For 39 of 41 events with high-confidence phase calls, the CNN-LOH or loss was inferred to produce homozygosity or hemizygosity of the inherited deletion, removing the reference (non-deletion) allele from the genome. (This dynamic resembles that of *ATM* in suggesting clonal selection for the rare, inherited risk allele.) The 70-kb deletion increased risk of 15q mosaicism by a factor of 698 (442–1,102): 45 of 89 carriers exhibited detectable 15q events (32 CNN-LOH, 2 loss, 11 ambiguous between CNN-LOH and loss). Notably, the 70-kb deletion was sometimes inherited on an allele that also had an independent 290-kb duplication of the locus (Extended Data Fig. 6); on this more complex allele, TM2D3 and TARSL2 gene dosage were normal. Carriers of the more complex allele did not exhibit predisposition to mCAs. Further study will be required to determine a proliferative mechanism involving TM2D3, TARSL2, or noncoding elements within the region.

The high penetrances (up to 50%) for the above *cis* associations led us to suspect that some risk-allele carriers might harbour multiple

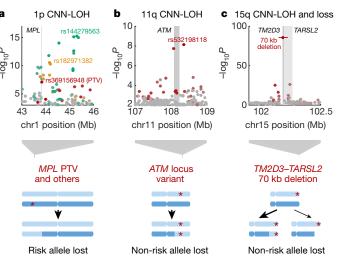


Fig. 4 | Novel loci associated with mCAs in *cis* due to clonal selection. a, *MPL*. b, *ATM*. c, *TM2D3–TARSL2*. In each locus, one or more inherited genetic variant predisposes chromosomal mutations to create a proliferative advantage. Bottom, genomic modifications; top, association *P* values (Fisher's exact test, n = 120,664 individuals). Independent lead associated variants are labelled, and variants are coloured according to linkage disequilibrium (LD) with lead variants (in shades of red, gold, or green; variants in grey are not in LD with lead variants). In c, the differing arrow weights to CNN-LOH and loss events indicate that CNN-LOH is the more common scenario (both in the population and among carriers of the risk variant).

subclonal cell populations with the associated alterations. Using a modified version of our methodology, we detected 39 individuals who had acquired two or more CNN-LOH mutations (with different breakpoints and allelic fractions) involving the same chromosome (Extended Data Fig. 8 and Supplementary Note 1). For all 39 individuals with multiple same-chromosome CNN-LOH events, all events involved recurrent selection of the same haplotype (in different clones). Of these 39 haplotypes, 16 carried a risk allele identified by our association scans, 13 appeared to involve other (undiscovered) alleles at the same loci, 5 duplicated 13q14 deletions, and 5 involved other genomic loci (Extended Data Fig. 8). This result indicates strong proliferative advantage conferred by CNN-LOH in these individuals and suggests that mitotic recombination occurs sufficiently frequently to yield multiple opportunities for clonal selection in individuals carrying inherited haplotypes with different proclivities for proliferation.

We also found two common variants on chromosome X that weakly increase risk of X loss while strongly influencing (in heterozygous females) which X chromosome is lost in the expanded clone. These involved a strong association ($P=6.6\times10^{-27}$, 1.9:1 bias in the lost haplotype) at Xp11.1 near DXZ1 and a weaker association ($P=1.0\times10^{-9}$, 1.5:1 bias in the lost haplotype) at Xq23 near DXZ4 (Table 1, Supplementary Table 9, and Supplementary Note 9). These associations do not appear to be explained by biased X chromosome inactivation³⁹ (Supplementary Table 10) and hint at yet another mechanism, different from those we have described.

Trans associations with mCAs

Genetic variants near genes involved in cell proliferation and cell cycle regulation predispose for male loss of $Y^{19,21}$, and female loss of X is also heritable ($h^2 = 26\%$ (17.4–36.2%) in sib-pair analysis)²¹, but no associations for X loss have previously been reported, to our knowledge. We confirmed the heritability of female X loss by performing BOLT-REML⁴⁰ analysis (see Methods), obtaining a SNP-heritability estimate of $h_g^2 = 10.6\%$ (s.e. 3.6%). Genome-wide association analysis for *trans* variants influencing X loss further revealed two genome-wide significant associations at the *SP140L* and HLA loci (Table 1).

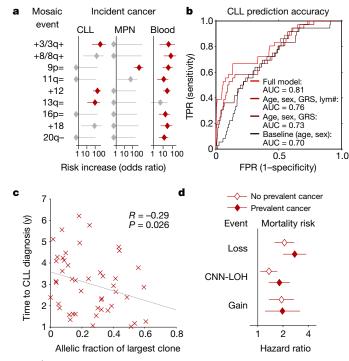


Fig. 5 | Associations between mCAs and incident cancers and mortality. a, Multiple mCA types confer increased risk of incident blood cancers diagnosed >1 year after DNA collection in n = 109,819 individuals with normal blood counts at assessment (Cochran-Mantel-Haenszel test adjusting for age and sex; error bars, 95% CI). b, A logistic model including mosaic status for 13q and trisomy 12 events along with other risk factors achieves high out-of-sample prediction accuracy for incident CLL (n = 36 cases and 113,923 controls with no cancer history). Lym#, log lymphocyte count. c, Time to malignancy tracks inversely with clonal cell fraction in n = 46 individuals with detectable clonality (of any mCA) who were diagnosed with CLL after assessment (one-sided Pearson's test). d, Loss, gain, and CNN-LOH events (on any autosome) all confer increased mortality risk in n = 128,854 individuals with no cancer history and n = 15,782 with prevalent cancers (error bars, 95% CI). Sample exclusions are detailed in the Methods. Numeric data are provided in Supplementary Tables 12 and 13.

Germline variants affecting cancer risk or chromosome-maintenance phenotypes could in principle increase the risk of clonal expansions. We considered 86 variants that have been implicated in previous genome-wide association studies (GWAS) on CLL, MPN, Y loss, clonal haematopoiesis, and telomere length and tested these variants for trans association with seven classes of mCAs, stratifying events by copy number and by autosome versus X chromosome. Four variants reached Bonferroni significance ($P < 8.3 \times 10^{-5}$): two linked variants in TERT^{11,20,41}, a rare frameshift mutation in CHEK2²⁰, and a lowfrequency 3' untranslated region (UTR) SNP in TP53^{21,42} (Supplementary Table 11). The TERT and CHEK2 variants associated with multiple types of autosomal event; by contrast, the TP53 SNP primarily associated with losses (both focal autosomal deletions and X losses). Carriers of the CHEK2 frameshift mutation were especially prone to developing multiple mCAs (one-sided binomial P = 0.008): 8 of 33 carriers with detected autosomal mosaicism had two or more mCAs, generally in multiple clones.

Mosaic chromosomal alterations and subsequent health

Cancer-free individuals with detectable mosaicism (at any locus) have a more than tenfold elevated risk of subsequent haematological cancer¹⁻⁴. For CLL, a slowly progressing cancer that is known to be preceded by clonal mosaicism years before progression^{43,44}, mosaic alterations observed in patients who go on to develop CLL occur at the same loci as those observed in patients with CLL^{32,33,45,46}. Using data on health outcomes for UK Biobank participants 4–9 years (median 5.7 years)

after DNA sampling, we identified nine specific mCAs that were significantly associated (FDR 0.05) with subsequent haematological cancer diagnoses (more than 1 year after DNA collection) in analyses corrected for age and sex and restricted to individuals with normal blood counts at assessment (Fig. 5a and Supplementary Table 12), confirming and providing additional resolution to previous findings^{1,2}. A logistic model combining mosaic status for CLL-associated events with other risk factors—age, sex, CLL genetic risk score⁴⁷, and lymphocyte count—achieved high CLL prediction accuracy (area under the curve (AUC) = 0.81) in tenfold cross-validation (Fig. 5b and Extended Data Fig. 9). Most of this predictive power came from early clones with trisomy 12, which we could detect at very low cell fractions (Extended Data Figs. 9, 10). Individuals with incident CLL exhibited clonality up to six years before diagnosis, and clonal fraction inversely correlated with time to malignancy (Fig. 5c). We further observed that detectable mosaicism roughly doubled risk for all-cause mortality (corrected for age, sex, and smoking status). This association was explained only partly by cancer deaths (Fig. 5d and Supplementary Table 13) and could reflect effects on cardiovascular illness¹², although further study is needed to explore this finding and rule out residual confounding.

Discussion

Mosaicism typically results from mutation followed by selective proliferation¹⁰, and our results uncover diverse biological mechanisms underlying this transformation. We identified very rare inherited variants that affect either the likelihood of mutation (at FRA10B) or its proliferative impacts (due to CNN-LOH in cis), and we also observed trans influences on clonal haematopoiesis in the cell cycle genes TP53, CHEK2, and TERT. Our findings of cis risk loci for CNN-LOH expansions are particularly noteworthy: while some CNN-LOH expansions have previously been observed to provide a second hit to a frequently mutated locus⁴⁸ or to disrupt imprinting⁴⁹, here we observed that CNN-LOHs can also achieve strong selective advantage by duplicating or removing inherited alleles. The high penetrances (up to 50%) of the inherited CNN-LOH risk variants we identified challenge what is usually seen as a fundamental distinction between inherited alleles and (more capricious) acquired mutations. A large fraction of carriers of the inherited alleles subsequently acquire and then clonally amplify the mutations in question. The high penetrances imply that mitotic recombination is sufficiently common to predictably unleash latent, inherited opportunities for clonal selection of homozygous cells during the lifespan of an individual, corroborating a recent observation of this phenomenon in skin⁵⁰. Similarly, we observed Mendelian inheritance patterns for 10q breakage at FRA 10B, despite this event involving an acquired mutation.

Clonal expansions exhibit varying levels of proliferation and biological transformation and thus have a spectrum of effects on health ¹⁰. We found that many mCAs, including some of those driven by *cis*-acting genetic variation, had no discernible adverse effects. However, mCAs commonly seen in blood cancers strongly increased cancer risk and could potentially be used for early detection—although we caution that these results are based on relatively short follow-up (4–9 years of cancer outcomes) and need independent replication. As population-scale efforts to collect genotype data and health outcomes continue to expand—increasing both sample sizes and the power of population-based chromosomal phasing—we anticipate ever-more-powerful analyses of clonal haematopoiesis and its clinical sequelae.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0321-x.

Received: 2 August 2017; Accepted: 16 May 2018; Published online 11 July 2018.

- Jacobs, K. B. et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. 44, 651-658 (2012).
- 2. Laurie, C. C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. 44, 642-650 (2012).
- 3. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. 371, 2477-2487 (2014).
- 4. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. 371, 2488-2498 (2014).
- 5 Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. 20, 1472-1478 (2014).
- McKerrell, T. et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Reports 10, 1239-1245 (2015).
- Machiela, M. J. et al. Characterization of large structural genetic mosaicism in human autosomes. Am. J. Hum. Genet. 96, 487-497 (2015).
- Vattathil, S. & Scheet, P. Extensive hidden genomic mosaicism revealed in normal tissue. Am. J. Hum. Genet. 98, 571-578 (2016).
- Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. É. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7**, 12484 (2016).
- 10. Forsberg, L. A., Gisselsson, D. & Dumanski, J. P. Mosaicism in health and disease - clones picking up speed. Nat. Rev. Genet. 18, 128–142 (2017).
- Zink, F. et al. Clonal hematopoiesis, with and without candidate driver
- mutations, is common in the elderly. *Blood* **130**, 742–752 (2017). Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**, 111–121 (2017).
- Acuna-Hidalgo, R. et al. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am. J. Hum. Genet. 101, 50-64 (2017).
- Laken, S. J. et al. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat. Genet. 17, 79-83 (1997).
- Jones, A. V. et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat. Genet. 41, 446-449 (2009).
- Kilpivaara, O. et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. Nat. Genet. 41, 455-459 (2009).
- 17. Olcaydu, D. et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. Nat. Genet. 41, 450-454 (2009).
- Koren, A. et al. Genetic variation in human DNA replication timing. Cell 159, 1015-1026 (2014).
- Zhou, W. et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. Nat. Genet. 48, 563-568 (2016).
- Hinds, D. A. et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood 128, 1121-1128
- Wright, D. J. et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. Nat. Genet. 49, 674-679 (2017)
- Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, PLoS Med. 12. e1001779 (2015).
- Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
- Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
- Machiela, M. J. et al. Mosaic chromosome 20q deletions are more frequent in the aging population. *Blood Advances* **1**, 380–385 (2017).
- Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron 68, 192-195 (2010).
- Werling, D. M. et al. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. Nat. Genet. 50, 727-736 (2018).
- Beroukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899-905 (2010).
- Davoli, T. et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell 155, 948-962
- Machiela, M. J. et al. Female chromosome X mosaicism is age-related and preferentially affects the inactivated X chromosome. Nat. Commun. 7, 11843
- Sinclair, E. J., Potter, A. M., Watmore, A. E., Fitchett, M. & Ross, F. Trisomy 15 associated with loss of the Y chromosome in bone marrow: a possible new aging effect. Cancer Genet. Cytogenet. 105, 20-23 (1998)
- Landau, D. A. et al. Mutations driving CLL and their evolution in progression and relapse. Nature 526, 525-530 (2015).
- Puente, X. S. et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
- Sutherland, G. R., Baker, E. & Seshadri, R. S. Heritable fragile sites on human chromosomes. V. A new class of fragile site requiring BrdU for expression. Am. J. Hum. Genet. **32**, 542–548 (1980).
- Hewett, D. R. et al. FRA10B structure reveals common elements in repeat expansion and chromosomal fragile site genesis. Mol. Cell 1, 773-781 (1998)
- Richards, R. I. & Sutherland, G. R. Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**, 709–712 (1992). Gurney, A. L., Carver-Moore, K., de Sauvage, F. J. & Moore, M. W.
- Thrombocytopenia in c-mpl-deficient mice. Science 265, 1445–1447 (1994).

- 38. Tefferi, A. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. Leukemia 24, 1128-1138 (2010).
- Tukiainen, T. et al. Landscape of X chromosome inactivation across human tissues. Nature 550, 244-248 (2017).
- 40. Loh, P.-R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat. Genet. 47, 1385–1392 (2015).
- 41. Oddsson, A. et al. The germline sequence variant rs2736100_C in TERT associates with myeloproliferative neoplasms. Leukemia 28, 1371-1374 (2014).
- 42. Stacey, S. N. et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat. Genet. 43, 1098-1103 (2011).
- Rawstron, A. C. et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N. Engl. J. Med. 359, 575-583 (2008).
- 44. Landgren, O. et al. B-cell clones as early markers for chronic lymphocytic leukemia. N. Engl. J. Med. 360, 659-667 (2009).
- Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell **152**, 714–726 (2013).
- Ojha, J. et al. Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. *Leukemia* **28**, 2395–2398 (2014).
- 47. Berndt, S. I. et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. Nat. Commun. 7, 10933 (2016).
- O'Keefe, C., McDevitt, M. A. & Maciejewski, J. P. Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. Blood 115, 2731-2739 (2010).
- 49. Chase, A. et al. Profound parental bias associated with chromosome 14 acquired uniparental disomy indicates targeting of an imprinted locus. Leukemia 29, 2069-2074 (2015).
- 50. Choate, K. A. et al. Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. Science 330, 94-97 (2010).

Acknowledgements We thank Y. Jakubek for assistance with follow-up on del(10q) events⁸ and G. Bhatia, A. Gusev, M. Lipson, X. Liu, L. O'Connor, N. Patterson, and B. van de Geijn for discussions. This research was conducted using the UK Biobank Resource under Application #19808. A.L.P. was supported by NIH grants R01 HG006399, R01 GM105857, R01 MH101244, and R21 HG009513. P.-R.L. was supported by NIH fellowship F32 HG007805, a Burroughs Wellcome Fund Career Award at the Scientific Interfaces, and the Next Generation Fund at the Broad Institute of MIT and Harvard. G.G., R.E.H., and S.A.M. were supported by NIH grant R01 HG006855 and the the Stanley Center for Psychiatric Research. H.K.F. was supported by the Fannie and John Hertz Foundation. Y.A.R. was supported by NIH award T32 GM007753, a National Defense Science and Engineering Graduate Fellowship, and the Paul and Daisy Soros Foundation. S.F.B. and G.G. were supported by US Department of Defense Breast Cancer Research Breakthrough Awards W81XWH-16-1-0315 and W81XWH-16-1-0316 (project BC151244). S.F.B. was supported by the Elsa U. Pardee Foundation and NCI MSKCC Cancer Center Core Grant P30 CA008748. M.E.T. was supported, in part, by NIH grants UM1 HG008900 and R01 HD081256. Computational analyses were performed on the Orchestra High Performance Compute Cluster at Harvard Medical School, which is partially supported by grant NCRR 1S10RR028832-01, and on the Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara and financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. This work was supported by a grant from the Simons Foundation (CEARLA was de #346040 and #346040). by a grant from the Simons Foundation (SFARI Awards #346042 and #385027, M.E.T.). We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren and E. Wijsman). We appreciate access to genetic and phenotypic data on SFARI Base.

Reviewer information Nature thanks S. Chanock, D. Conrad, I. Hall and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions P.-R.L., G.G., S.F.B., S.A.M., and A.L.P. designed the study. P.-R.L. and G.G. analysed UK Biobank data. R.E.H. analysed SSC data. P.-R.L., G.G., H.K.F., and Y.A.R. developed statistical methods. P.F.P. assisted with IBD analyses. B.M.B. assisted with cancer phenotype curation. M.E.T. and S.A.M. supervised SSC analyses. All authors wrote the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0321-x.

Supplementary information is available for this paper at https://doi.org/ 10.1038/s41586-018-0321-x.

Reprints and permissions information is available at http://www.nature.com/ reprints.

Correspondence and requests for materials should be addressed to P-R.L. or G.G. or S.A.M. or A.L.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

UK Biobank cohort and genotyping intensity data. The UK Biobank is a very large prospective study of individuals aged 40–70 years at assessment ²². Participants attended assessment centres between 2006 and 2010, where they contributed blood samples for genotyping and blood analysis and answered questionnaires about medical history and environmental exposures. In the years since assessment, health outcome data for these individuals (for example, cancer diagnoses and deaths) have been accrued via UK national registries.

We analysed genetic data from the UK Biobank interim release (about 30% of the full UK Biobank) consisting of 152,729 samples typed on the Affymetrix UK BiLEVE and UK Biobank Axiom arrays with about 800,000 SNPs each and more than 95% overlap. We removed 480 individuals marked for exclusion from genomic analyses based on missingness and heterozygosity filters and one individual who had withdrawn consent, leaving 152,248 samples. We restricted the variant set to biallelic variants with missingness $\leq 10\%$ and we further excluded 111 variants found to have significantly different allele frequencies between the UK BiLEVE array and the UK Biobank array, leaving 725,664 variants on autosomes and the X chromosome. Finally, we additionally excluded 118,139 variants for which fewer than 10 samples (or for chrX, fewer than 5 female samples) were called as homozygous for the minor allele; we observed that genotype calls at these variants were susceptible to errors in which rare homozgyotes were called as heterozygotes. We phased the remaining 607,525 variants using Eagle2 24 with –Kpbwt = 40,000 and otherwise default parameters.

We transformed genotyping intensities to $\log_2 R$ ratio (LRR) and B-allele frequency (BAF) values⁵¹ (which measure total and relative allelic intensities) after affine-normalization and GC wave-correction⁵² in a manner similar to that described¹ (Supplementary Note 1). For each sample, we then computed s.d. (BAF) among heterozygous sites within each autosome, and we removed 320 samples with median s.d. (BAF) > 0.11 indicating low genotyping quality. Finally, we removed an additional 725 samples with evidence of possible contamination⁸ (based on apparent short interstitial CNN-LOH events in regions of long-range linkage disequilibrium; Supplementary Note 1) and one sample without phenotype data, leaving 151,202 samples for analysis.

Detection of mCAs using long-range haplotype phase. Here we outline the key ideas of our approach to mCA detection; full details are provided in Supplementary Note 1. The core intuition is to harness long-range phase information to search for local imbalances between maternal and paternal allelic fractions in a cell population (Extended Data Fig. 1). The utility of haplotype phase for this purpose has previously been recognized^{8,53,54}, but previous approaches have needed to account for phase switch errors occurring roughly every megabase, a general challenge faced by haplotype-based analyses⁵⁵. In the UK Biobank, we have phase information accurate at the scale of tens of megabases^{23,24}, enabling a new modelling approach and considerable gains in sensitivity for detection of large events at low cell fractions (Supplementary Note 5). (Because our method is phase-based, it has the limitation that it cannot detect events contained within regions of homozygosity. While this issue is minor in our study of large events, other approaches originally developed for detection of shorter constitutional or high-cell-fraction CNVs are not subject to this limitation^{56,57}.)

Our technique employs a three-state hidden Markov model (HMM) to capture mCA-induced deviations in allelic balance ($|\Delta BAF|$) at heterozygous sites. (By contrast, the hapLOH method 8,54 tabulates 'switch consistency' between consecutive heterozygous sites.) Our model has a single parameter Θ , which represents the expected absolute BAF deviation at germline hets within an mCA. In computationally phased genotyping intensity data, multiplying phase calls with (signed) BAF deviations produces contiguous regions within the mCA in which the expected phased BAF deviation is either $+\Theta$ or $-\Theta$ (with sign flips at phase switch errors); outside the mCA, no BAF deviation is expected. The three states of our HMM encode these three possibilities, and emissions from the states represent noisy BAF measurements. Transitions between the $+\Theta$ and $-\Theta$ states represent switch errors, while transitions between \pm Θ and the 0 state capture mCA boundaries.

Modelling observed phased BAF deviations using a parameterized HMM has the key benefit of naturally producing a likelihood ratio test statistic for determining whether a chromosome contains a mCA. Explicitly, for a given choice of Θ , we can compute the total probability of the observed BAF data under the assumption that mCA-induced BAF deviations have $E[|\Delta BAF|] = \Theta$, using standard HMM dynamic programming computations to integrate over uncertainty in phase switches and mCA boundaries. Taking the ratio of the maximum likelihood over all possible choices of Θ to the likelihood for $\Theta = 0$ (that is, no mCA) yields a test statistic. If the HMM perfectly represented the data, this test statistic could be compared to an asymptotic distribution. However, we know in practice that parameters within the HMM (for example, transition probabilities) are imperfectly

estimated, so we instead calibrated our test statistic empirically: we estimated its null distribution by computing test statistics on data with randomized phase, and we used this empirical null to control FDR. Finally, for chromosomes passing the FDR threshold, we called mCA boundaries by sampling state paths from the HMM (using the maximum likelihood value of Θ).

The above detection procedure uses only BAF data and ignores LRR measurements by design (to be maximally robust to genotyping artefacts); however, after detecting events, we incorporated LRR data to call detected mCAs as loss, CNN-LOH, or gain. All mosaic chromosomal alterations cause BAF (measuring relative allelic intensity) to deviate from 0.5 at heterozygous sites, and losses and gains cause LRR (measuring total intensity) to deviate from 0, with deviations increasing with clonal cell fraction; accordingly, we observed that plotting detected events by LRR and BAF deviation produced three linear clusters (Fig. 2a), consistent with previous work $^{\rm l.2.8}$. We called copy number using chromosome-specific clusters to take advantage of the differing frequencies of event types on different chromosomes. Because the clusters converge as BAF deviation approaches zero, we left copy number uncalled for detected mCAs at low cell fraction (with <95% confident copy number), comprising 29% of all detected mCAs. We then estimated clonal cell fractions as described $^{\rm l}$.

As a post-processing step to exclude possible constitutional duplications, we filtered events of length $>\!10$ Mb with LRR $>\!0.35$ or with LRR $>\!0.2$ and $|\Delta {\rm BAF}|\!>\!0.16$, and we filtered events of length $<\!10$ Mb with LRR $>\!0.2$ or with LRR $>\!0.1$ and $|\Delta {\rm BAF}|\!>\!0.1$. We chose these thresholds conservatively based on visual inspection of LRR and BAF distributions, in which likely constitutional duplications formed well-defined clusters (Supplementary Note 1). (Most constitutional duplications were already masked in a pre-processing step involving a separate HMM described in Supplementary Note 1.)

Enrichment of mCA types in blood lineages. We analysed 14 blood count indices (counts and percentages of lymphocytes, basophils, monocytes, neutrophils, red cells, and platelets, as well as distribution widths of red cells and platelets) from complete blood count data available for 97% of participants. We restricted the analysis to individuals of self-reported European ancestry (96% of the cohort), leaving 140,250 individuals; we then stratified by sex and quantile-normalized each blood index after regressing out age, age squared, and smoking status.

To identify classes of mCAs linked to different blood cell types, we first classified mCAs based on chromosomal location and copy number. For each autosome, we defined five disjoint categories of mCAs that comprised the majority of detected events: loss on p arm, loss on q arm, CNN-LOH on p arm, CNN-LOH on q arm, and gain. We subdivided loss and CNN-LOH events by arm but did not subdivide gain events because most gain events are whole-chromosome trisomies (Fig. 1). For chromosome X, we replaced the two loss categories with a single whole-chromosome loss category. Altogether, this classification resulted in 114 mCA types. We restricted our blood cell enrichment analyses to 78 mCA types with at least 10 occurrences, and we further excluded the chr17 gain category (because nearly all of these events arise from i(17q) isochromosomes already counted as 17p– events; Supplementary Note 2).

For each of the 77 remaining mCA types, we computed enrichment of mCAs among individuals with anomalous (top 1%) values of each normalized blood index using Fisher's exact test (two-sided; P values reported throughout this manuscript are from two-sided statistical tests unless explicitly stated otherwise). We reported significant enrichments passing an FDR threshold of 0.05 (Fig. 2f and Supplementary Table 6).

Chromosome-wide association tests for cis associations with mCAs. To identify inherited variants influencing nearby mCAs, we performed two types of association analysis. First, we searched for variants that increased the probability of developing nearby mCAs. For each variant, we performed a Fisher test for association between the variant and up to three variant-specific case-control phenotypes, defined by considering samples to be cases if they contained loss, CNN-LOH, or gain events containing the variant or within 4 Mb (to allow for uncertainty in event boundaries). We tested phenotypes with at least 25 cases; in total, 48 out of $69 = 23 \times 3$ possible event types had at least 25 carriers, and the rest were excluded from association analyses. We performed these tests on 51 million imputed variants with minor allele frequency (MAF) $> 2 \times 10^{-5}$ (imputed by UK Biobank using merged UK10K and 1000 Genomes Phase 3 reference panels⁵⁸), excluding variants with non-European MAF greater than five times their European MAF, which tended to be poorly imputed. We analysed 120,664 individuals who remained after restricting to individuals of self-reported British or Irish ancestry, removing principal component outliers (>4 s.d.), and imposing a relatedness cutoff of 0.05 (using plink --rel-cutoff 0.05)⁵⁹. (In our non-GWAS analyses, which focused on mosaic individuals, we did not apply any special handling of related individuals as the number of related pairs was very small: for example, only 11 third-degree or closer relationships among 4,889 individuals with autosomal mosaicism.)

We also ran a second form of association analysis searching for variants for which mCAs tended to shift allelic balance (analogous to allele-specific

expression). For a given class of mCAs, for each variant, we examined heterozygous mosaic individuals for which the mCA overlapped the variant, and we performed a binomial test to check whether the mCA was more likely to delete or duplicate one allele rather than the other. We restricted the binomial test to individuals in which the variant was confidently phased relative to the mCA (that is, no disagreement in five random resamples from the HMM used to call the mCA).

Given that the two association tests described above are independent, we applied a two-stage approach to identify robust genome-wide significant associations. We used a P value threshold of 10^{-8} for discovery in either test and then checked for nominal P < 0.05 significance in the other test (reasoning that variants that influenced mCAs would exhibit both types of association). At all loci with $P < 10^{-8}$ for either test, the most significant variant with $P < 10^{-8}$ in one test reached nominal significance in the other (Table 1). At identified loci, we further searched for secondary independent associations reaching $P < 10^{-6}$.

In our final analyses, we refined mCA phenotypes to slightly increase power to map associations. For the loci associated with 1p, 9p, and 15q CNN-LOH, we found that association strength improved by expanding case status to include all events reaching the telomere (because several detected telomeric events with uncertain copy number were probably actually CNN-LOH events associated with the same germline variants). For the association signal at *FRA10B*, we refined case status to only include terminal loss events extending from 10q25 to the telomere (because of the breakpoint specificity of this event). We verified that all association tests produced well-calibrated test statistics (Supplementary Note 3).

Identity-by-descent analysis at *MPL* **and** *FRA10B.* **At** loci for which we found evidence of multiple causal rare variants, we searched for long haplotypes shared identical-by-descent among mCA carriers to further explore the possibility of additional or recurrent causal variants. We called IBD tracts using GERMLINE with haplotype extension⁶⁰.

Simons Simplex Collection WGS data set. The Simons Simplex Collection (SSC) is a repository of genetic samples from autism simplex families collected by the Simons Foundation Autism Research Initiative (SFARI) 26 . We analysed 2,079 whole-genome sequences from the first phase of SSC sequencing (median coverage $37.8 \times)^{27}$ to examine whether mCAs we detected contributed to genetic risk of autism. (The main data set consisted of 2,076 individuals in 519 quartets; we additionally analysed three individuals that did not belong to a complete quartet but were of interest based on high read counts at *FRA10B*.)

Detection and calling of 70-kb deletion at 15q26.3. We discovered the inherited 70-kb deletion associated with 15q CNN-LOH and loss by mapping the 15q26.3 association signal (specifically, the rs182643535 tag SNP) in WGS data (Fig. 4c and Extended Data Fig. 6). We then called this deletion in the UK Biobank SNP-array data using genotyping intensities at 24 probes in the deleted region (Extended Data Fig. 7).

Detection and imputation of VNTRs at *FRA10B*. For all WGS samples with 10 or more reads at the *FRA10B* locus, we attempted to perform local assembly of the reads and identify a primary VNTR motif in the assembly. We identified 12 distinct primary motifs carried by 26 individuals in 14 families (Extended Data Fig. 5a, b and Supplementary Note 8). Owing to read dropout in many samples, it is possible that these VNTR motifs may be found in additional samples, and that other VNTR motifs may not have been detected. We imputed the VNTR sequences into UK Biobank using Minimac3⁶¹. Full details are provided in Supplementary Note 8.

GWAS and heritability estimation for *trans* drivers of clonality. We tested variants with MAF > 1% for *trans* associations with six classes of mCAs (any event, any loss, any CNN-LOH, any gain, any autosomal event, any autosomal loss) on 120,664 unrelated individuals with European ancestry (described above) using BOLT-LMM⁶², including 10 principal components, age, and genotyping array as covariates. We also tested association with female X loss using an expanded set of 3,462 likely X loss calls at an FDR of 0.1, restricting this analysis to 66,685 female individuals. In our targeted analysis of 86 variants implicated in previous GWAS, we applied a Bonferroni significance threshold of 8.3×10^{-5} based on 86 variants and 7 phenotypes. We estimated SNP heritability of X loss using BOLT-REML⁴⁰, transforming estimates to the liability scale⁶³.

Analysis of X chromosome inactivation in GEUVADIS RNA sequencing data. To test for possible mediation of preferential X haplotype loss by biased X chromosome inactivation (XCI), we examined GEUVADIS RNA sequencing (RNA-seq) data⁶⁴ for evidence of biased XCI near the primary biased loss association at Xp11.1. We identified three coding SNPs in *FAAH2* within the pericentromeric linkage disequilibrium block containing the association signal. We analysed RNA-seq data for 61 European-ancestry individuals who were heterozygous for at least one SNP (60 of 61 were heterozygous for all three SNPs, and the remaining individual was heterozygous at two of the SNPs). We used GATK ASEReadCounter⁶⁵ to identify allele-specific expression from RNA-seq BAM files. Most individuals displayed strong consistent allele-specific expression across the three SNPs, as expected for XCI in clonal lymphoblastoid cell lines³⁹; however, we observed

no evidence of systematically biased XCI in favour of one allele or the other (Supplementary Table 10).

UK Biobank cancer phenotypes. We analysed UK cancer registry data provided by UK Biobank for 23,901 individuals with one or more prevalent or incident cancer diagnoses. Cancer registry data included date of diagnosis and ICD-O-3 histology and behaviour codes, which we used to identify individuals with diagnoses of CLL, MPN, or any blood cancer^{66,67}. Because our focus was on prognostic power of mCAs for predicting diagnoses of incident cancers more than one year after DNA collection, we excluded all individuals with cancers reported prior to this time (either from cancer registry data or self-report of prevalent cancers). We also restricted our attention to the first diagnosis of cancer in each individual, and we censored diagnoses after 30 September 2014, as suggested by UK Biobank (resulting in a median follow-up time of 5.7 years, s.d. 0.8 years, range 4–9 years). Finally, we restricted analyses to individuals with self-reported European ancestry. These exclusions reduced the total counts of incident cases to 78 (CLL), 42 (MPN), and 441 (any blood cancer), which we analysed with 119,330 controls. In our primary analyses, we further eliminated individuals with any evidence of potential undiagnosed blood cancer based on anomalous blood counts (lymphocyte count outside the normal range of $1-3.5 \times 10^9$ /l, red cell count >6.1 × 10^{12} /l for males or $>5.4 \times 10^{12}$ /l for females, platelet count $>450 \times 10^9$ /l, red cell distribution width >15%), leaving incident case counts of 36 (CLL), 23 (MPN), and 327 (any blood cancer).

Estimation of cancer risk conferred by mCAs. To identify classes of mCAs associated with incident cancer diagnoses, we classified mCAs based on chromosomal location and copy number into the 114 classes described above. We then restricted our attention to the 45 classes with at least 30 carriers (to reduce our multiple hypothesis burden, given that we would be underpowered to detect associations with the rarer events). For each mCA class, we considered a sample to be a case if it contained only the mCA or if the mCA had the highest cell fraction among all mCAs detected in the sample (that is, we did not count carriers of subclonal events as cases). We computed odds ratios and P values for association between mCA classes and incident cancers using Cochran-Mantel-Haenszel (CMH) tests to stratify by sex and by age (in six 5-year bins). We used the CMH test to compute odds ratios (for incident cancer any time during follow-up) rather than using a Cox proportional hazards model to compute hazard ratios because both the mCA phenotypes and the incident cancer phenotypes were rare, violating normal approximations underlying regression. We reported significant associations passing an FDR threshold of 0.05 (Fig. 5a and Supplementary Table 12).

Prediction of incident CLL. We considered four nested logistic models for prediction of incident CLL. In the first model, a baseline, we included only age and sex as explanatory variables. In the second model, we added CLL genetic risk (computed using 14 high-confidence GWAS hits that had both been previously published⁴⁷ and reached $P < 5 \times 10^{-8}$). In the third model, we added log lymphocyte count. In the full model, we added explanatory variables for 13q and +12 events.

We assessed the accuracy of each model on two benchmark sets of samples. We restricted our primary analyses to individuals with normal lymphocyte counts $(1-3.5\times10^9/l)$ at assessment (that is, exhibiting at most slight clonality); in auxiliary analyses, we removed this restriction (and expanded the full prediction model to include 11q-, +12, 13q-, 13q CNN-LOH, 14q-, 22q-, and the total number of other autosomal events). We performed tenfold stratified cross-validation to compare model performance. We assessed prediction accuracy by merging results from all cross-validation folds and computing area under the receiver operating characteristic curve (AUC) (Fig. 5b), and we also measured precision-recall performance (Extended Data Fig. 9). (We caution that while AUC is commonly used to assess discriminative power, AUC does not have a direct clinical interpretation⁶⁸.) Estimation of mortality risk conferred by mCAs. We analysed UK death registry data provided by UK Biobank for 4,619 individuals reported to have died since assessment. We censored deaths after 31 December 2015, as suggested by UK Biobank, leaving 4,518 reported deaths over a median follow-up time of 6.9 years (range 5-10 years). We examined the relationship between mCAs and mortality, aiming to extend previous observations that mosaic point mutations increase mortality risk^{3,4,11}. For this analysis, we were insufficiently powered to stratify mCAs by chromosome owing to the weaker effects of mCAs on mortality risk and the relatively small number of deaths reported during follow-up. We therefore stratified mCAs only by copy number and computed the hazard ratio conferred by each event class using a Cox proportional hazards model. We restricted these analyses to individuals with self-reported European ancestry, and we adjusted for age and sex as well as smoking status, which was previously associated with clonal haematopoiesis^{3,11,69} and associates with mosaicism in the UK Biobank (P=0.00017). We did not exclude individuals based on blood counts in these analyses (or in our time-to-malignancy versus clonal fraction analyses), hence the larger sample sizes in Fig. 5c, d than in Fig. 5a, b.

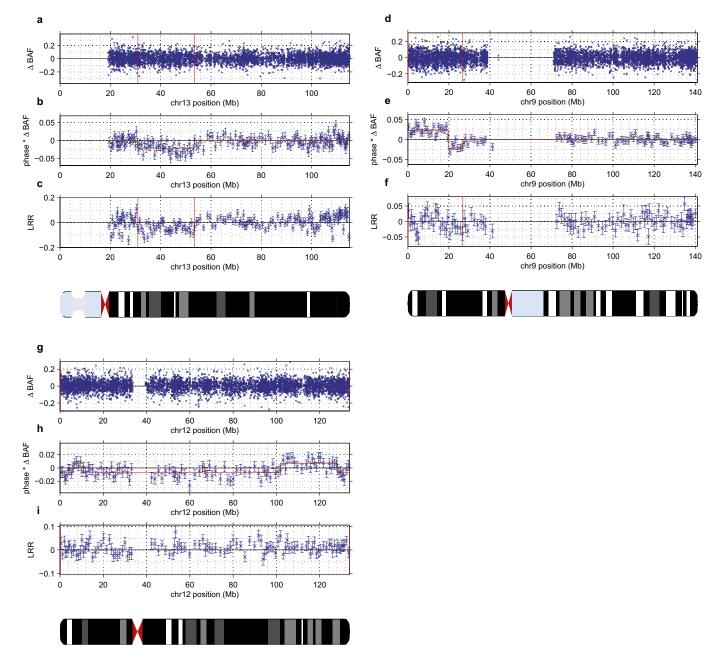
 $\label{lem:code} \textbf{Code availability.} \ Code \ used \ to perform \ the \ analyses \ in \ this \ study \ is \ available \ from \ the \ corresponding \ authors \ upon \ request.$



Data availability. Mosaic event calls are available in the Supplementary Data. Access to the UK Biobank Resource is available via application (http://www.ukbiobank.ac.uk/). Approved researchers can obtain the SSC population data set described in this study by applying at https://base.sfari.org.

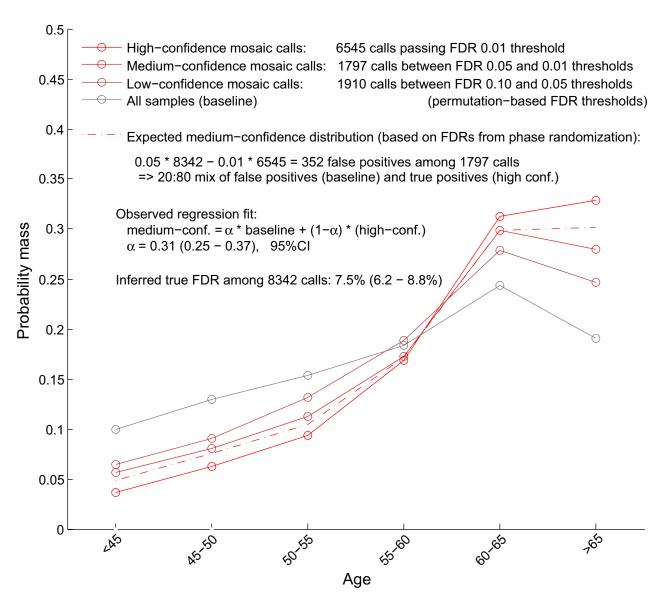
- Peiffer, D. A. et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 16, 1136–1148 (2006).
- Diskin, S. J. et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 36, e126 (2008).
- Nik-Zainal, S. et al. The life history of 21 breast cancers. Cell 149, 994–1007 (2012).
- Vattathil, S. & Scheet, P. Haplotype-based profiling of subtle allelic imbalance with SNP arrays. Genome Res. 23, 152–158 (2013).
- Genovese, G., Leibon, G., Pollak, M. R. & Rockmore, D. N. Improved IBD detection using incomplete haplotype information. *BMC Genet.* 11, 58 (2010).
- Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572 (2004).
- Pique-Regi, R., Cáceres, À. & Gónzález, J. R. R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics* 11, 380 (2010).
- Huang, J. et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat. Commun. 6, 8111 (2015).

- Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).
- Gusev, A. et al. Whole population, genome-wide mapping of hidden relatedness. Genome Res. 19, 318–326 (2009).
- Das, S. et al. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287 (2016).
- 62. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
- Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 88, 294–305 (2011).
- Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511 (2013).
- Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195 (2015).
- Turner, J. J. et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood* 116, e90–e98 (2010).
- Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127, 2391–2405 (2016).
- Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* 17, 392–406 (2016).
- 69. Dumanski, J. P. et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science **347**, 81–83 (2015).



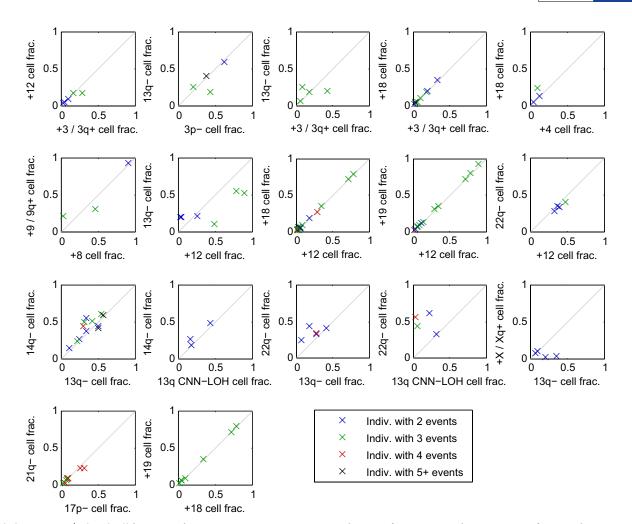
Extended Data Fig. 1 | Examples of mosaic events called using phased genotyping intensities. \mathbf{a} - \mathbf{c} , UK Biobank mCA sample 2791 has a mosaic deletion of chr13 from approximately 31–53 Mb that cannot be confidently called from unphased BAF and LRR data (\mathbf{a} , \mathbf{c}). However, the existence of an event is evident in the phased BAF data (\mathbf{b}), and the regional decrease in LRR indicates that this event is a deletion. In \mathbf{b} , mean phased BAF is plotted for SNPs aggregated into bins spanning n = 25 heterozygous sites; the same bins are used for \mathbf{c} . Error bars, s.e.m. \mathbf{d} - \mathbf{f} , Sample 1645 has a mosaic CNN-LOH on chr9p from the 9p telomere to about 26 Mb that cannot be confidently called from unphased BAF data (\mathbf{d}) but is evident in phased BAF data (\mathbf{e}). A phase switch error causes a sign flip in phased

BAF at approximately 20 Mb. The lack of a shift in LRR in the region (f) indicates that this event is a CNN-LOH. In $\bf e$, mean phased BAF is plotted for SNPs aggregated into bins spanning n=50 heterozygous sites; the same bins are used for $\bf f$. Error bars, s.e.m. $\bf g-i$, Sample 2464 has a full-chromosome mosaic event on chr12 that cannot be confidently called from unphased BAF and LRR data ($\bf g$, $\bf i$) but is evident in phased BAF data ($\bf h$). Several phase switch errors cause sign flips in phased BAF across chr12. The slight positive shift in mean LRR ($\bf i$) indicates that this event is most likely to be a mosaic gain of chr12. In $\bf h$, mean phased BAF is plotted for SNPs aggregated into bins spanning n=50 heterozygous sites; the same bins are used for $\bf i$. Error bars, s.e.m.



Extended Data Fig. 2 | Estimation of true FDR using age distributions of individuals with mCA calls. We generated age distributions for (i) 'high-confidence' detected events passing a permutation-based FDR threshold of 0.01 (bright red); (ii) 'medium-confidence' events below the FDR threshold of 0.01 but passing an FDR threshold of 0.05 (darker red); and (iii) 'low-confidence' events below the FDR threshold of 0.05 but passing an FDR threshold of 0.10 (darkest red; not analysed but plotted for context). We compared these distributions to the overall age distribution of UK Biobank participants (grey). On the basis of the numbers of events in each category, approximately 20% of medium-confidence detected

events are expected to be false positives. To estimate our true FDR, we regressed the medium-confidence age distribution on the high-confidence and overall age distributions, reasoning that the medium-confidence age distribution should be a mixture of correctly called events with age distribution similar to that of the high-confidence events, and spurious calls with age distribution similar to the overall cohort. We observed a regression weight of 0.31 for the component corresponding to spurious calls, in good agreement with expectation, and implying a true FDR of 7.5% (6.2–8.8%, 95% CI based on regression fit on n=6 age bins).



Extended Data Fig. 3 | Clonal cell fractions of co-occurring events generally suggest co-existence within the same cell population. For each pair of significantly co-occurring events (Fig. 2b), we compared the clonal fractions of the two events within each individual that carried both events. Each point in the plots corresponds to an individual carrying the pair of events under consideration; individuals are colour-coded by the total number of events they carry. For nearly all pairs of events, the clonal fractions of the two events were very similar in most individuals carrying both events, suggesting that the events occurred in the same clonal cell

population. A few exceptions do seem to exist; for example, 22q- versus 13q CNN-LOH cell fraction; here, the cell fractions suggest that 13q CNN-LOH events may be present in a subclone. This observation is consistent with acquired uniparental disomy of 13q providing a second hit within a del(13q14) clonal expansion, as we see in Extended Data Fig. 8. (We did not include del(13q14) vs. 13q CNN-LOH in this plot because inference of clonal fractions is complex for these overlapping events; see Extended Data Fig. 8.)

Observed here: Known mechanism: Jones et al. 2009 Nat Genet, JAK2 46/1 Kilpivaara et al. 2009 Nat Genet, haplotype Olcaydu et al. 2009 Nat Genet, Koren et al. 2014 Cell JAK2 46/1 predisposes to JAK2 V617F mutation (untyped here) 15 JAK2 Subsequent CNN-LOH d^{01} bol-(detectable here) 10 confers further proliferative advantage 5 => 9p CNN-LOH associates with 46/1 4.5 5 5.5

Extended Data Fig. 4 | Replication of previous association between JAK2 46/1 haplotype and 9p CNN-LOH in cis due to clonal selection. The common JAK2 46/1 haplotype has previously been shown to confer risk of somatic JAK2 V617F mutation such that subsequent 9p CNN-LOH produces a strong proliferative advantage^{15–18,20} (right). In our analysis, CNN-LOH on 9p is strongly associated with JAK2 46/1 ($P=1.6 \times 10^{-13}$, OR = 2.7 (2.1–3.5); Fisher's exact test on n=120,664 individuals) with

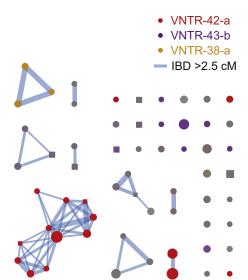
chr9 position (Mb)

the risk haplotype predominantly duplicated by CNN-LOH in hets (52 of $n\!=\!61$ heterozygous cases; binomial $P\!=\!1.8\times10^{-8}$). Left, the genomic modification is illustrated in the top panel and association signals are plotted in the bottom. The lead associated variant is labelled, and variants are coloured according to linkage disequilibrium with the lead variant (scaled for readability).

a Variable Number Tandem Repeats (VNTRs) at FRA10B identified in WGS data

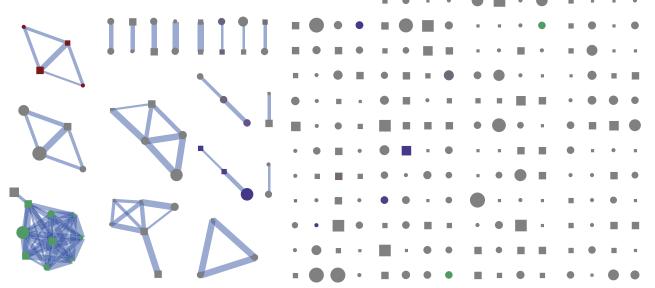
VNTR-43-e	GATATAATATAT	CGTATACAA	TATATACGT	ATATAT	TATATAC
E13	GATATAATATAT	CATATGTAA	TAGATAT-G	ATGTAT	TACATAT
E11	GATATAATATAT	CATATGTAA	TAGATATGT	ATATAT	TATATAT
E12	GATATAATATAT	CATATGTAA	TAGATATGT	ATATAT	TACATAT
E8	GATATAATATAT	ATTATATAA	TATATATGT	ATATAT	TATATAT
VNTR-43-a	GATATAATATAT	CGTATATAT	TATATACGG	ATACAT	TATATAT
HG19-REF	GATATAATATAT	ACATAT	TATATATGT	ATATAT	TATATAT
VNTR-42-c	GATATAATATAT	CATACATAT	TATATAT-G	ATATAT	TATCTAT
VNTR-42-b	GATATAATATAT	CATACATAT	TATCTAT-G	ATATAT	TATATAT
VNTR-42-d	GATATAATATAT	CATACATAT	TATGTAT-G	ATATAT	TATATAT
VNTR-42-e	GATATCATATAT	CATACATAT	TATATAT-G	ATATAT	TATATAT
VNTR-39-a	GATATAATATAT	CCATAT	TATATAT-G	ATATAT	TATATAT
VNTR-43-b	GATATAATATAT	CATACATAT	TATATATGG	ATATAT	TATATAT
VNTR-42-a	GATATAATATAT	CATACATAT	TATATAT-G	ATATAT	TATATAT
VNTR-38-a	GATATAATATAT	CATAT	TATATAC-G	ATATAT	GATATAT
E10	GATATAATATAT	CATATATAA	TATATATGT	ATATAT	TATATAT
VNTR-43-d	GATATAATATAT	CATATATAA	TATATATGG	ATATAT	TATATAT
E17	GATATAATATAT	CATATATAA	TATATAT-G	ATATAT	TATATAT
VNTR-43-c	GATATAATATAT	CATATATAT	TATATACGG	ATATAT	TATATAT
E19	GATATAATATAT	CATATATAT	TATATAT-G	ATATAT	TATATAT
	*****	*	** **	** **	* **

b Identity-by-descent graph at FRA10B for UK Biobank del(10q) individuals colored according to imputed VNTRs



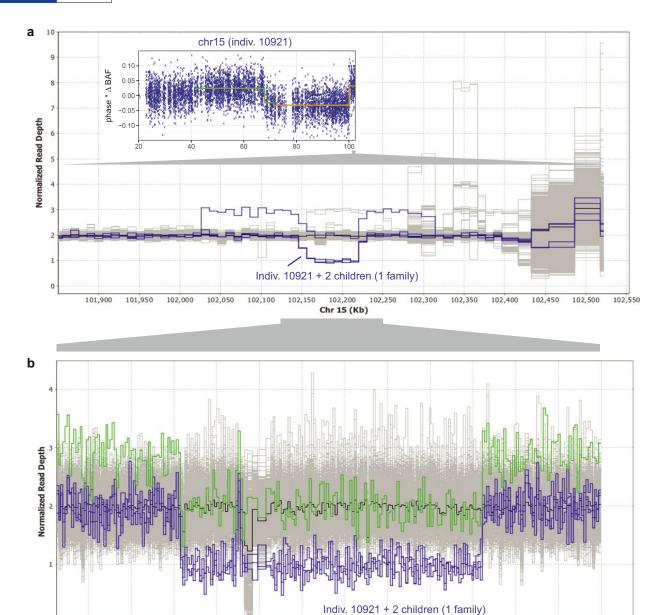
c IBD graph at MPL for UK Biobank 1p mosaic individuals

Edges = IBD>2.5cM (edge weights increase with IBD length)
Red nodes = carriers of rare MPL nonsense mutation (rs369156948)
Green nodes = carriers of long rare haplotype (tag: rs144279563)
Blue nodes = carriers of long rare haplotype (tag: rs182971382)



Extended Data Fig. 5 | Evidence of multiple causal variants for 10q25.2 breakage and 1p CNN-LOH associations. a, Multiple expanded repeats at *FRA10B* drive breakage at 10q25.2. We identified 12 distinct primary repeat motifs at *FRA10B* in 26 whole-genome-sequenced individuals from 14 families (labelled VNTR-*N*-*x*, where *N* denotes length in base pairs); carriers of these repeats exhibit varying degrees of *FRA10B* repeat expansion (Supplementary Note 8). The repeat motifs are AT-rich and are similar to *FRA10B* repeats previously reported³⁵. The alignment provided here includes the repeat motifs that were most frequently observed in *FRA10B* expanded alleles³⁵ (E8, E13, E17, and E19) along with a few other closely related expanded repeat motifs (E10, E11, and E12). b, Carriers

of the 10q terminal deletion in the UK Biobank share long haplotypes at 10q25.2 identical-by-descent. Square nodes in the IBD graph correspond to males and circles to females. Node size is proportional to cell fraction and edge weight increases with IBD length. Coloured nodes indicate imputed carriers of variable number tandem repeats (VNTRs) at *FRA10B* (Supplementary Table 7); colour intensity scales with imputed dosage. c, Identity-by-descent graph at *MPL* locus (chr1:43.8 Mb) on individuals with mCAs on chr1 extending to the p telomere. Colored nodes indicate imputed carriers of SNPs independently associated with mosaic 1p CNN-LOH (Fig. 4a).



Extended Data Fig. 6 | Germline CNVs at 15q26.3. a, Read depth profile plot of WGS samples in the terminal 700 kb of chr15q. Three individuals in one family carry an approximately 70-kb deletion at 15q26.3, and a fourth carries the same deletion along with an approximately 290-kb duplication (probably on the same haplotype, based on population frequencies of these events; see Extended Data Fig. 7). These four individuals (highlighted in blue) segregate with the rs182643535:T allele in the WGS cohort. Inset: the parental carrier in the family, individual 10921,

102,150

102,160

102,170

102,180

102,190

Chr 15 (Kb)

102,200

102,210

102,220

102,140

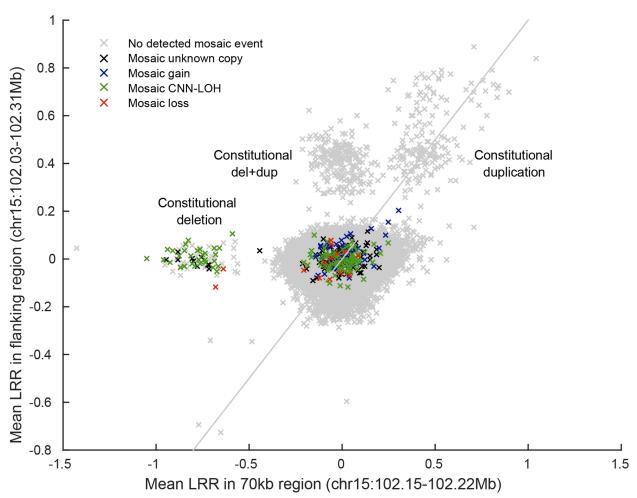
102,130

has detectable mosaicism in two distinct 15q CNN-LOH subclones (one starting at 41.64 Mb with 4.6% cell fraction, the other starting at 71.64 Mb with an additional 2.0% cell fraction). **b**, Expanded read depth profile plot, with deletion-only individuals highlighted in blue and the del \pm dup individual highlighted in green. Breakpoint analysis indicates that the deletion spans chr15:102151467–102222161 and contains a 1,139-bp mid-segment (chr15:102164897–102166035) that is retained in inverted orientation. The duplication spans chr15:102026997–102314016.

102,230

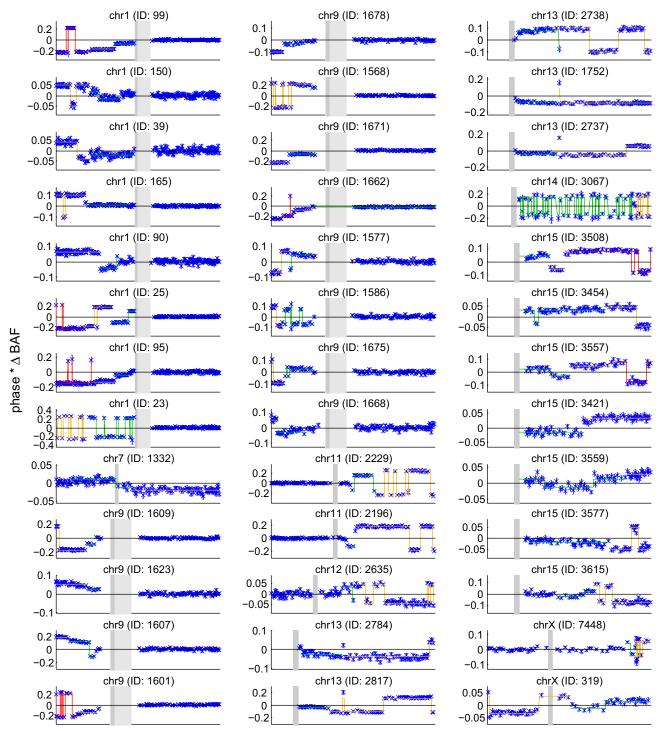
102,240

102,250



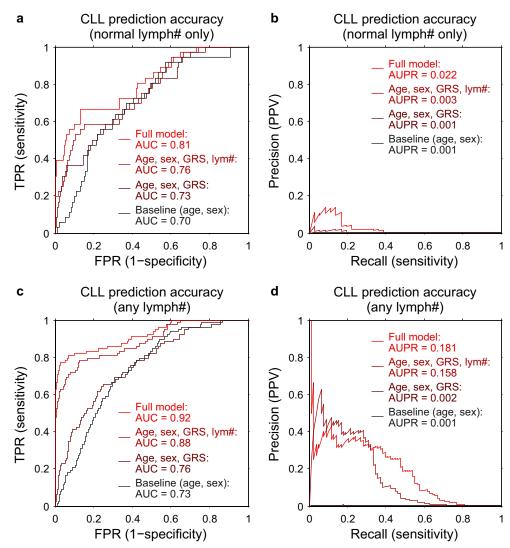
Extended Data Fig. 7 | Mosaic chromosomal alterations and germline CNVs at 15q26.3. Using identified breakpoints of the germline 70-kb deletion and 290-kb duplication (Extended Data Fig. 6), we computed mean genotyping intensity (LRR) in UK Biobank samples within the 70-kb deletion region (24 probes) and within the flanking 220-kb region (97 probes). Individuals are plotted by flanking 220-kb mean LRR versus 70-kb mean LRR and coloured according to mosaic status for somatic 15q mCAs. UK Biobank samples carrying the 70-kb deletion, 290-kb

duplication, and both (del+dup) are all easily identifiable in distinct clusters. The plot also appears to contain clusters with higher copy number. Of the three CNV-carrying alleles, the simple 70-kb deletion is the only one that predisposes to mCAs. Most mosaic events containing the 70-kb deletion are CNN-LOH events that make cells homozygous for the 70-kb deletion; two individuals have somatic loss of the homologous (normal) chromosome, making cells hemizygous for the 70-kb deletion.



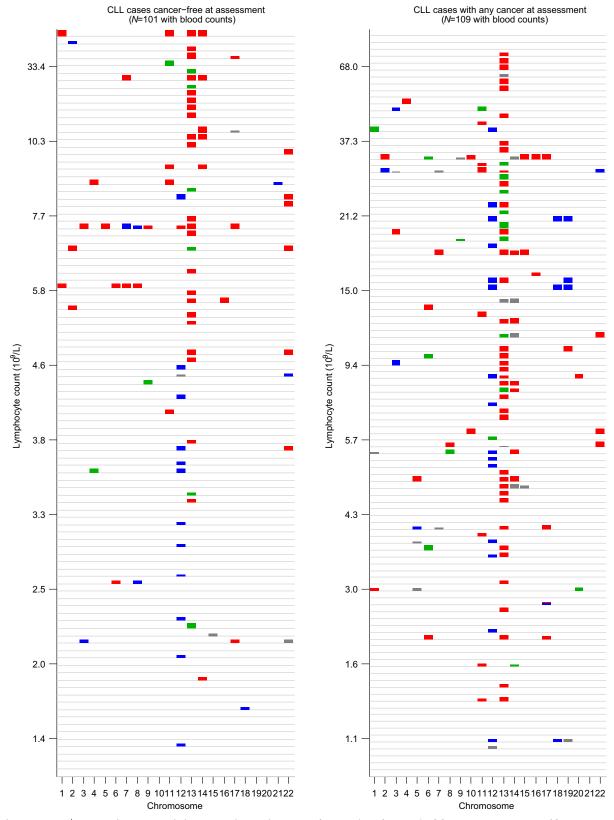
Extended Data Fig. 8 | Phased BAF plots of chromosomes with multiple CNN-LOH subclones. All of the plots exhibit step functions of increasing $|\Delta \text{BAF}|$ towards a telomere, which is the hallmark of multiple clonal cell populations containing distinct CNN-LOH events that affect different spans of a chromosomal arm (all extending to the telomere). Distinct $|\Delta \text{BAF}|$ values (called using an HMM) are indicated with different colours. Flips in the sign of phased BAF usually correspond to phase

switch errors. Two samples exhibit high switch error rates: 14q individual 3067 (explained by non-European ancestry), and 1p individual 23 (explained by very high $|\Delta {\rm BAF}|$; extreme shifts in genotyping intensities result in poor genotyping quality). All five individuals with multiple CNN-LOH events on chr13q appear to contain switch errors over 13q14, but these switches are actually explained by overlapping 13q14 deletions; see Supplementary Note 1 for detailed discussion.



Extended Data Fig. 9 | CLL prediction accuracy: receiver operating curves and precision-recall curves. CLL prediction benchmarks using tenfold stratified cross validation on: only individuals with lymphocyte counts in the normal range ($1 \times 10^9/L$ to $3.5 \times 10^9/L$), as in our primary analyses (n=36 cases, 113,923 controls) ($\bf a, \bf b$); and individuals with any lymphocyte count (n=78 cases, 118,481 controls) ($\bf c, \bf d$). a matches Fig. 5b, and $\bf b$ shows the precision-recall curve from the same analysis. $\bf c$ and $\bf d$ correspond to an analogous analysis in which we removed the restriction

on lymphocyte count and also used additional mosaic event variables for prediction (11q-, 14q-, 22q-, and total number of autosomal events). In both benchmarks, individuals with previous cancer diagnoses or CLL diagnoses within 1 year of assessment were excluded; however, some individuals with very high lymphocyte counts pass this filter (and probably already had CLL at assessment despite being undiagnosed for more than 1 year), hence the difference in apparent prediction accuracy between the two benchmarks.



Extended Data Fig. 10 \mid Mosaic chromosomal alterations detected in CLL cases sorted by lymphocyte count. Individuals are stratified by cancer status at DNA collection (no previous diagnosis versus any previous

diagnosis), and mCAs (red, loss; green, CNN-LOH; blue, gain; grey, undetermined) are plotted per chromosome as coloured rectangles (with height increasing with BAF deviation).



Single-cell analysis of early progenitor cells that build coronary arteries

Tianying Su^{1,12}, Geoff Stanley^{2,12}, Rahul Sinha^{3,12}, Gaetano D'Amato¹, Soumya Das¹, Siyeon Rhee¹, Andrew H. Chang¹, Aruna Poduri¹, Brian Raftrey¹, Thanh Theresa Dinh^{4,5}, Walter A. Roper^{4,5}, Guang Li⁶, Kelsey E. Quinn⁷, Kathleen M. Caron⁷, Sean Wu^{3,6,8}, Lucile Miquerol⁹, Eugene C. Butcher^{4,5}, Irving Weissman³, Stephen Quake^{10,11} & Kristy Red-Horse¹*

Arteries and veins are specified by antagonistic transcriptional programs. However, during development and regeneration, new arteries can arise from pre-existing veins through a poorly understood process of cell fate conversion. Here, using single-cell RNA sequencing and mouse genetics, we show that vein cells of the developing heart undergo an early cell fate switch to create a pre-artery population that subsequently builds coronary arteries. Vein cells underwent a gradual and simultaneous switch from venous to arterial fate before a subset of cells crossed a transcriptional threshold into the pre-artery state. Before the onset of coronary blood flow, pre-artery cells appeared in the immature vessel plexus, expressed mature artery markers, and decreased cell cycling. The vein-specifying transcription factor COUP-TF2 (also known as NR2F2) prevented plexus cells from overcoming the pre-artery threshold by inducing cell cycle genes. Thus, vein-derived coronary arteries are built by pre-artery cells that can differentiate independently of blood flow upon the release of inhibition mediated by COUP-TF2 and cell cycle factors.

The ability of cells to switch fates and acquire new identities is critical for organogenesis and regeneration, but the mechanisms that underlie cell fate conversions are poorly understood. The vasculature is a model for this process because it initially differentiates into arteries and veins whose transcriptional networks antagonize each other (Notch signalling maintains arteries while COUP-TF2 maintains veins^{1,2}). However, during development and regeneration, veins can become the source of new arteries³⁻⁶. The timing and requirements of vein-to-artery conversions are not known, but could inform artery regeneration.

In mice, a portion of the coronary arteries of the heart develop from a vein called the sinus venosus (SV; Fig. 1a). During embryogenesis, endothelial cell-lined angiogenic sprouts migrate from the SV to fill the heart with an immature coronary vessel plexus⁴. This plexus unites with plexus vessels from the endocardium^{4,7,8}, and, together, they remodel into arteries, capillaries and veins. The plexus lacks blood flow until it attaches to the aorta, and arterial morphogenesis requires this event, suggesting that blood flow initiates artery development^{8–11}. However, it has been difficult to delineate cell fate changes during coronary angiogenesis owing to the limited number of molecular markers and bulk transcriptional analyses of heterogeneous populations.

Single-cell RNA sequencing (scRNA-seq) can overcome this limitation by producing single-cell-resolution maps of developmental transitions. Here, we developed a statistical test that categorizes subpopulations within scRNA-seq data sets as continuous or discrete to identify candidate developmental transitions. Computational or in vivo analysis of the SV-to-coronary transition revealed that SV cells of the mouse heart undergo a gradual conversion from vein to artery before a subset crosses a threshold to differentiate into pre-artery cells. Pre-artery cells differentiated before blood flow from the SV and endocardium and produced a large portion of coronary arteries. COUP-TF2 blocked progression to the pre-artery state through

activation of cell cycle genes, which ultimately inhibited artery development. Understanding this and other cell fate switches and inhibitory signals will advance our knowledge of tissue development and could improve regenerative medicine.

Finding developmental transitions in scRNA-seq data

We performed a two-step analysis that identified and clustered cell subtypes by iterative robust principal component analysis (rPCA), and then subjected clusters to a pairwise discreteness test (Fig. 1b). First, cell subtype clusters were manually defined on the basis of unique gene expression patterns and cell separation in multiple iterations of rPCA¹² (Fig. 1b). rPCA was better than classical PCA at separating small subpopulations of cells¹³ (Extended Data Fig. 1a). We also replaced default principal component scores with a sum of the top 60 genes score because it was less correlated with technical artefact and better correlated with cluster-specific genes (Extended Data Fig. 1b, c). Cell cycle heterogeneity was also removed (Extended Data Fig. 1d), and plots were inspected to confirm the absence of doublets (Extended Data Fig. 1e). This process resulted in cell clusters that correlate well with genes that define cell identity, and not with cell cycle heterogeneity or technical artefact (Extended Data Fig. 1c, d).

Second, we developed the pairwise discreteness test to determine whether clusters were discrete or continuous (that is, connected by intermediate or transitioning cells). This statistical test projects pairs of subpopulations onto a linear axis of cell identity, measures the size of the gap between the populations, and estimates the number of intermediate cells (Fig. 1b and Extended Data Fig. 1f). It also determines the strength of continuity (Extended Data Fig. 1h), and could be confirmed using simulated data (Extended Data Fig. 1h). Combining the results created a relationship graph (Fig. 1b), which could identify candidate developmental transitions. Then, cell fate changes could be analysed

¹Department of Biology, Stanford University, Stanford, CA, USA. ²Program in Biophysics, Stanford University, Stanford, CA, USA. ³Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. ⁴Veterans Affairs Palo Alto Health Care System and The Palo Alto Veterans Institute for Research, Palo Alto, CA, USA. ⁵Department of Pathology, Stanford University, Stanford, CA, USA. ⁶Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁷Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NCA, USA. ⁸Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ⁹Aix-Marseille Université, CNRS UMR 7288, IBDM, Marseille, France. ¹⁰Department of Bioengineering, Stanford University, Stanford, CA, USA. ¹¹Chan Zuckerberg Biohub, San Francisco, CA, USA. ¹²These authors contributed equally: Tianying Su. Geoff Stanley, Rahul Sinha. *e-mail: kredhors@stanford.edu

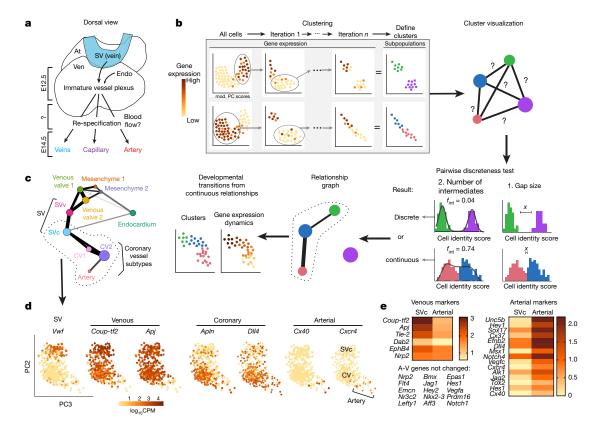


Fig. 1 | **Identifying pre-artery cells using scRNA-seq. a**, **b**, Schematics of coronary artery development (**a**) and computational pipeline (f_{inb} estimated fraction of cells that are intermediate; x, width of the largest gap in scores between populations) (**b**). **c**, Relationship graph for

ApjCreER-traced endothelial subtypes. **d**, Pre-artery cells extend from the plexus in the SVc–coronary vessel continuum. Gene expression in brown. n = 415 cells. **e**, Heat map of venous and arterial genes. At, atria; endo, endocardium; ven, ventricle.

in high resolution by observing gene expression changes across continuous populations (Fig. 1b).

We used this pipeline to analyse 843 *ApjCreER* lineage-labelled (Cre expressed in SV) cardiac endothelial cells from hearts removed from mouse embryos at embryonic day 12.5 (E12.5) (Extended Data Fig. 1g). Our data set contained endothelial cells from the SV, SV-derived coronary vessels, venous valves, valve mesenchyme, and some ventricular endocardial cells (Extended Data Fig. 1i, j). Clustering and the pairwise discreteness test revealed a continuum between coronary vessel subtypes, the SV, venous valves, ventricular endocardium, and mesenchyme (*Pdgfra*⁺, *Pecam1*^{low/-}) (Fig. 1c and Extended Data Fig. 1i-k). These associations are consistent with anatomical relationships (SV is adjacent to venous valves and endocardium) and previous lineage tracing experiments (SV transitions into coronary vessels and endocardium transitions into mesenchyme)^{8,11,14-17}. Thus, our pipeline can identify subpopulations and recapitulate known developmental transitions and anatomical relationships.

Pre-artery cells differentiate before blood flow

We analysed the developmental transition linking SV coronary progenitors (SVc) and coronary vessels (Fig. 1c, dotted line). Only the SVc was included because clustering indicated that the SV had two domains (Fig. 1c), and this was confirmed using immunofluorescence and in situ hybridization (Extended Data Fig. 2a–f). The SVc was anatomically and transcriptionally continuous with coronary vessels, whereas the SVv (SV valve proximal) was continuous with venous valves (Fig. 1c, Extended Data Fig. 2d, f). Therefore, rPCA of the SVc and coronary vessels was performed to study the SVc–coronary vessel continuum (Fig. 1d).

Unexpectedly, the SVc-coronary vessel continuum identified cells that were transcriptionally distinct and expressed genetic markers of mature arterial cells (Fig. 1d). We previously reported⁴ that plexus cells express arterial genes such as *Dll4* and *Efnb2*, but these are also

expressed in angiogenic vessels, and are not artery-specific 18,19 . The scRNA-seq analysis revealed that, within the $Dll4^+$ domain, some cells had initiated a distinctive transcriptional program, shifting away in the rPCA plot (Fig. 1d). Cells within this subset specifically expressed mature artery-specific genes, including Cx40 (also known as Gja5) (Fig. 1d). Analysis of multiple arterial and venous genes in single cells or as averages within clusters (defined in Extended Data Fig. 2g) revealed that many arterial genes were either specific to or significantly increased in the $Cx40^+$ cluster (Fig. 1d and Extended Data Fig. 3a, b). Multiple venous genes were either completely depleted or significantly downregulated (Fig. 1d and Extended Data Fig. 3c, d). Comparison of expression between the SVc and arterial populations revealed that SV-derived cells showed an extensive switch towards arterial fate (Fig. 1e).

We next compared E12.5 arterial cells with adult coronary vessel cells. Each embryonic cell was matched to the adult cell to which it was most similar within the artery–capillary–vein continuum formed by adult coronary vessels²⁰ (Extended Data Fig. 3e–g). E12.5 artery cells were most similar to adult arterial cells, whereas coronary vessel plexus cells were most similar to adult capillaries and veins (Extended Data Fig. 3h). We also found that E12.5 and adult artery populations were enriched for nearly the same artery markers (Supplementary Table 1). The exception was *Notch1* (enriched only in adults), possibly because blood flow upregulates *Notch1*²¹, and E12.5 is before the onset of coronary perfusion. Thus, a subpopulation of plexus cells undergo a transcriptional shift to resemble mature arteries before the presence of arterial vessels or blood flow, prompting us to term them pre-artery cells.

The scRNA-seq also identified new arterial genes (Extended Data Fig. 4a). *Slc45a4* marked pre-artery cells at early stages and was later specific to mature embryonic arteries (Extended Data Fig. 4b, c). It was also enriched in adult coronary artery cells (Extended Data Fig. 4a). We found other genes to be enriched in pre-artery cells (Extended

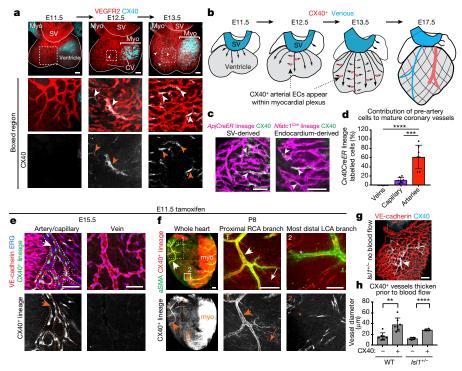


Fig. 2 | Pre-artery cells build coronary arteries. a, CX40 immunofluorescence in hearts to mark pre-artery cells (arrowheads). b, Schematic of pre-artery cells during coronary development. c, CX40⁺ cells in SV- and endocardium-derived plexus. d, Cx40CreER lineage labelling (E11.5 induction). n=7 hearts. e, f, Pre-artery lineage labelling in arteries (arrowheads) and a subset of capillaries (arrows) at E15.5 (e)

Data Fig. 4d). Of these, *Mecom* and *Igfbp3* marked arteries in adults (Extended Data Fig. 4d).

Location, origins, and fate of pre-artery cells

In late embryonic stages (E17.5), CX40 is specific to mature arteries (Extended Data Fig. 5a). By contrast, whole-mount immunostaining at early stages revealed that a small population of CX40⁺ cells first appeared at E12.5. These cells were interspersed within the intramy-ocardial plexus and expanded by E13.5 (Fig. 2a, b). Localization of additional pre-artery genes confirmed this result (Extended Data Fig. 5b). The absence of CX40⁺ cells in the SV and their presence in the coronary vessel plexus agreed with clustering and pairwise analysis showing that pre-artery cells were continuous only with the coronary vessel plexus (Fig. 1c). Defining clusters using Seurat showed similar results (Extended Data Fig. 6a), although clusters were not as precise and were associated with cell cycle genes (Extended Data Fig. 6b, c). Thus, coronary angiogenesis involves the specification of single arterial endothelial cells within the intramyocardial plexus (Fig. 1b).

Although our scRNA-seq investigated only SV-derived vessels, lineage tracing revealed that coronary arteries are derived from both the SV and the endocardium 22 (Extended Data Fig. 6d). Single CX40 $^{+}$ cells were detected in the plexus from both sources (Fig. 2c), indicating that pre-artery specification occurs during both SV and endocardium angiogenesis.

Cx40CreER Rosa^{tdTomato} embryos were used for lineage tracing of pre-artery cells (tamoxifen, E11.5; Extended Data Fig. 6e, f). CX40⁺ pre-artery cells were later found in arteries, but not veins (Fig. 2d, e). A few capillaries were lineage-traced, indicating that pre-artery cells could revert to a capillary fate (Fig. 2d). Dosing at E10.5 ensured that our result was not due to persistent tamoxifen (Extended Data Fig. 6g), and clonal level labelling confirmed the lineage data (Extended Data Fig. 6h). Notably, at postnatal day (P)8, the right and left coronary artery branches were heavily lineage-labelled in hearts from mice dosed

and P8 (**f**). Asterisks in **f** indicate non-lineage labelled tips. **g**, CX40⁺ cells (arrowhead) in hearts that lack coronary blood flow. **h**, CX40⁺ vessels begin remodelling without blood flow. Wild-type, n=7 hearts; $Isl1^{+/-}$, n=6 hearts. **d**, **h**, Data shown as mean \pm s.d. myo, myocardium. Scale bars: **c**, 50 μ m; **a**, **e**, **f**, 100 μ m. Unpaired two-tailed t test. ** $P \le 0.01$; ***P < 0.001; ****P < 0.001.

at E11.5; only the most distal tips were unlabelled (Fig. 2f and Extended Data Fig. 6i). Thus, pre-artery cells build a large portion of mature coronary arteries.

Pre-artery cells first appeared before blood flow, but they were abundant in the plexus through E14.5 (Extended Data Fig. 7c), suggesting that specification could continue after coronary perfusion. To investigate this possibility, we used Cre lines that specifically label either coronary vessel plexus (*ApjCreER*) or pre-artery (*Cx40CreER*) cells (Extended Data Fig. 7a) and induced labelling at various times (Extended Data Fig. 7b). Labelling of the coronary vessel plexus at E12.5 or E13.5 lineage-traced a small number of pre-artery cells (Extended Data Fig. 7d, e). However, when the coronary vessel plexus was labelled at E14.5, there was no tracing into artery main branches and very little in the tips (Extended Data Fig. 7f, h). Conversely, labelling at E14.5 with Cx40CreER lineage-traced most left and right coronary artery branches (Extended Data Fig. 7g, h). Finally, inducing labelling with Cx40CreER at E16.5 resulted in few capillary cells being labelled at embryonic and postnatal stages (Extended Data Figs. 6j, 7i). These data indicate that pre-artery specification occurs in the coronary plexus between E12.5 and E14.5, creating a progenitor pool that forms virtually all of the embryonic left and right coronary artery branches.

We next investigated whether the artery tips that did not form from pre-artery cells (Fig. 2f) arose from pre-existing arteries or through capillary differentiation. Induction of *ApjCreER* and *Cx40CreER* labelling at P2 revealed that artery tips at P6 were composed of *ApjCreER*-lineage cells but depleted of *Cx40CreER*-labelled cells (Extended Data Fig. 6k). Thus, postnatal artery tips grow by capillary arterialization.

The morphogenic changes that accompany coronary artery remodelling are seen after blood flow has been established, and are thought to be triggered by shear stress^{9,10}. In E13.5 *Isl1* mutant mice that have delayed blood flow²³, pre-artery cells had congregated in the region where the coronary artery would eventually form (Fig. 2g) and began to increase

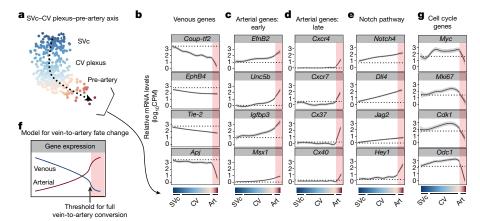


Fig. 3 | The venous-to-arterial fate change is gradual and culminates in an expression threshold. a, Coronary differentiation pathway (dashed arrow). b-e, Gene expression along the differentiation pathway. Dotted

lines, SVc expression levels; red shading, pre-artery cells. f, Model based on known marker gene patterns. g, Cell cycle genes decreased in pre-artery cells. Art, pre-artery; CV, coronary vessel.

lumen size (Fig. 2h). Therefore, pre-artery cells within the plexus can differentiate and initiate remodelling before cues from blood flow.

Gradual cell fate conversion

To investigate the vein-to-artery conversion, single cells along the SVc-coronary vessel plexus-pre-artery developmental transition were projected onto a linear continuum (Fig. 3a, b). Gene expression was then visualized by LOESS regression (Fig. 3b-e, g and Extended Data Fig. 8a). There was a progressive decrease in venous identity as cells exited the SV and moved towards pre-artery (see *Coup-tf2*, *EphB4* and *Tie-2* (also known as *Tek*); Fig. 3b). A sharp decrease in venous genes was seen in cells that had undergone full pre-artery specification (see *Coup-tf2*

and *Apj* (also known as *Aplnr*); Fig. 3b). Arterial gene expression showed two patterns: 'early' genes, expression of which progressively increased in coronary vessel plexus and pre-artery cells (Fig. 3c and Extended Data Fig. 8a), and 'late' genes, expression of which was low in coronary vessel plexus, but increased sharply in pre-artery cells (Fig. 3d and Extended Data Fig. 8a). Notch ligands and receptors were early genes, with the exception of *Hey1*, which increased sharply in pre-artery cells (Fig. 3e and Extended Data Fig. 8a). These findings suggest that the loss of venous identity is initially gradual with a progressive increase in arterial identity, and that pre-artery specification occurs after a threshold of venous loss and arterial gain has been achieved (Fig. 3f).

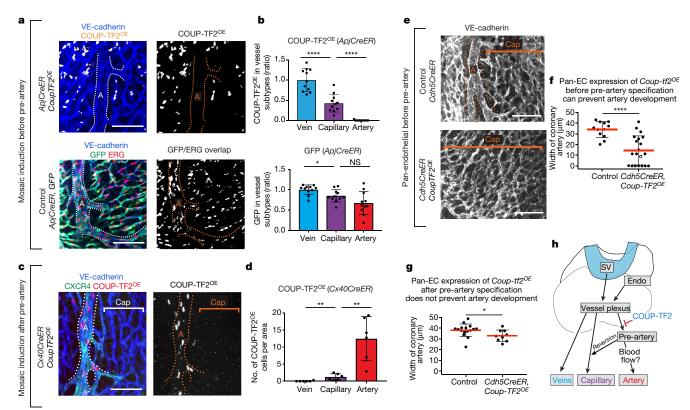


Fig. 4 | **COUP-TF2 specifically blocks pre-artery specification. a**, **b**, E15.5 hearts induced to express Coup- $tf2^{OE}$ or Gfp before pre-artery specification. **c**, **d**, E15.5 hearts induced to express Coup- $tf2^{OE}$ after pre-artery specification. **b**, Coup- $tf2^{OE}$, n = 11 hearts; Gfp, n = 11 hearts. **d**, n = 6 hearts. **e**, **f**, Coup- $tf2^{OE}$ induction in all endothelial cells before pre-artery specification. Control, n = 12 hearts; Coup- $tf2^{OE}$, n = 20 hearts.

g, Coup-tf2^{OE} induction in all endothelial cells after pre-artery specification. Control, n=16 hearts; Coup-tf2^{OE}, n=9 hearts. h, Schematic displaying differentiation step blocked by COUP-TF2. A, artery; Cap, capillary; Endo, endocardium. Scale bars, $100 \mu m$. Data shown as mean \pm s.d. P value: unpaired two-tailed t-test. NS, P>0.05; * $P\le0.05$; * $P\le0.01$; ****P<0.01.

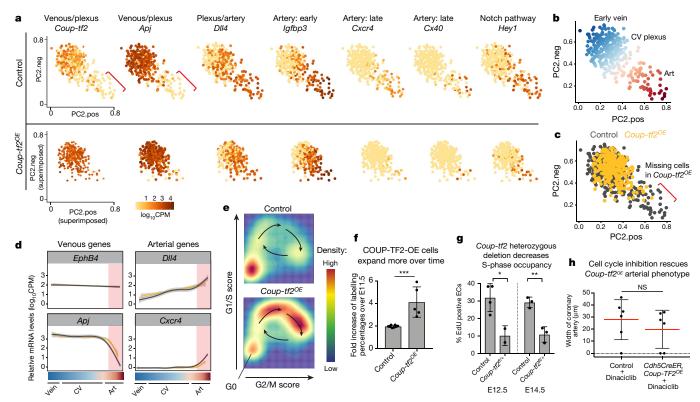


Fig. 5 | COUP-TF2 inhibits artery specification through cell cycle genes. a, rPCA plots from E14.5 hearts (wild-type, n=347 cells; Coup- $tf2^{OE}$, n=321 cells). Red brackets, artery cells devoid of Coup-tf2 and Apj. b, Coronary continuum based on gene expression patterns in a. n=347 cells. c, Coup- $tf2^{OE}$ cells do not populate the Coup- $tf2^{OE}$ artery population. Wild-type, n=347 cells; Coup- $tf2^{OE}$, n=321 cells. d, Progression towards artery is not generally affected by Coup- $tf2^{OE}$. Wild-type, grey lines; Coup- $tf2^{OE}$, yellow lines. Red-shaded region, pre-artery cells. e, Heat map showing the distribution of coronary vessel

plexus cells in the indicated cell cycle phases. **f**, Fold increase in control GFP or COUP-TF2^{OE} cells between E11.5 and E14.5. Control, n=8 hearts; $Coup-tf2^{OE}$, n=5 hearts. **g**, EdU incorporation in coronary endothelial cells from Cdh5CreER $Coup-tf2^{fl/+}$ hearts. E12.5: control, n=4 hearts; $Coup-tf2^{fl/+}$, n=2 hearts. E14.5: control, n=3 hearts; $Coup-tf2^{fl/+}$, n=3 hearts. **h**, Cell cycle inhibition reverses the ability of $Coup-tf2^{OE}$ to block artery formation (compare to Fig. 4f). Control, n=6 hearts; $Coup-tf2^{OE}$, n=6 hearts. P=0.4167. Data shown as mean \pm s.d. Unpaired two-tailed t test. $*P \le 0.05$; $**P \le 0.01$; $***P \le 0.001$.

To understand the pre-artery threshold, we performed pathway analysis using gene set enrichment analysis (GSEA)²⁴. Most pathways that were enriched in plexus over arterial cells were associated with cell cycling (Extended Data Fig. 8b). Arterial cells are thought to leave the cell cycle in response to blood flow^{25–27}; however, pre-artery cells collected before blood flow displayed a decrease in cell cycle genes (Fig. 3g, Extended Data Fig. 8c, and Supplementary Table 2). In vivo, pre-artery cells were less proliferative than the surrounding plexus (Extended Data Fig. 8d). Thus, decreased proliferation in arteries is acquired during pre-artery specification, and not specifically in response to blood flow.

COUP-TF2 blocks artery formation

To investigate whether pre-artery specification was necessary for artery formation, we required a tool to block this process. We tested COUP-TF2 because it induces venous fate and antagonizes arterial fate^{1,28} and was sharply decreased in pre-artery cells (Figs. 1d, 3b). *ApjCreER* mice were crossed to mice that constitutively express *Coup-tf2* after Cre recombination²⁹ (Extended Data Fig. 9a) and pregnant dams were treated with tamoxifen to induce overexpression of *Coup-tf2* (*Coup-tf2*^{OE}) before pre-artery specification. Cre recombination of the *Coup-tf2*^{OE} allele was low, making this experiment a mosaic analysis in which *Coup-tf* cells were followed within wild-type tissue (Extended Data Fig. 9b, c).

Coup-tf2^{OE} cells were present in capillaries and veins, but not arteries (Fig. 4a, b, top and Extended Data Fig. 9d). By contrast, control GFP⁺ cells were found in arteries, capillaries, and veins (Fig. 4a, b, bottom). Coup-tf2^{OE} cells could survive in arteries when VE-cadherin-CreER induced recombination after arteries had formed (Extended Data Fig. 9e). Coup-tf2^{OE} cells could also migrate normally onto the heart

(Extended Data Fig. 9f-h), although they caused a mild increase in vessel density at E13.5 (Extended Data Fig. 9g). Thus, forced COUP-TF2 expression before pre-artery specification blocks cells from contributing to coronary arteries, suggesting a failure to acquire pre-artery fate.

Induction of Coup- $tf2^{OE}$ after pre-artery specification with Cx40CreER(tamoxifen at E11.5 or E12.5) resulted in numerous Coup-tf2^{OE} cells within the artery (Fig. 4c, d, and Extended Data Fig. 9i) that expressed the arterial markers CXCR4 and JAG1 (Fig. 4c and Extended Data Fig. 9i). Therefore, Coup-tf2^{OE} inhibits arterial fate only before pre-artery specification. Pre-artery specification was then blocked throughout the entire coronary plexus by inducing widespread Coup*tf2*^{OE} recombination using *Cdh5CreER* (tamoxifen at E11.5 and E13.5). This resulted in small or completely absent coronary arteries (Fig. 4e, f). By contrast, induction of *Cdh5CreER-Coup-tf2*^{OE} after pre-artery specification, but before arterial morphogenesis (tamoxifen at E13.5 and E15.5), resulted in relatively normal artery development, confirming that the later steps in artery formation are not greatly inhibited by COUP-TF2 (Fig. 4g). Thus, pre-artery specification is required for artery development, and this is the specific differentiation step that is antagonized by COUP-TF2 (Fig. 4h).

COUP-TF2 inhibits pre-artery via cell cycle genes

We next used scRNA-seq to compare control and *Coup-tf2^{OE}* cells. E14.5 coronary endothelial cells (Extended Data Fig. 10a) were analysed as described for E12.5. *Coup-tf2^{OE}* cells were identified by the expression of the transgene's FLAG-myc tag (Extended Data Fig. 10b, c). rPCA revealed a transcriptional continuum linking venous, coronary vessel plexus, and arterial cells (Fig. 5a, b). Vein cells in this data set expressed *Coup-tf2* and *Apj* and lacked *Dll4* and *Notch4*, as has been

described for coronary veins^{4,30}. Superimposing transgenic cells onto the control continuum showed that Coup- $tf2^{OE}$ cells were excluded only from the arterial population (Fig. 5c). Venous and arterial genes along the continuum were not generally inhibited by Coup- $tf2^{OE}$ (Fig. 5a, d and Extended Data Fig. 10d). The defect instead was in the number of fully pre-artery or arterial cells, as shown with genes such as Cxcr4 and Cx40 (Fig. 5a).

Analysis of differential gene expression did not reveal marked changes in the expression of Notch genes, despite the prevailing theory that COUP-TF2 functions by antagonizing this pathway (Fig. 5a, d, Extended Data Fig. 10d and Supplementary Table 3). Furthermore, overexpression of Notch signalling did not rescue the *Coup-tf2^{OE}* phenotype (Extended Data Fig. 9j, k). It is possible that expression levels were not high enough to overcome COUP-TF2. Instead, a prominent feature of *Coup-tf2^{OE}* cells was an increase in cell cycle gene expression (Supplementary Table 3). Plotting coronary vessel plexus and vein cells according to G1/S/G2/M cell cycle staging revealed that the *Coup-tf2^{OE}* population contained more cells with a cycling profile when compared to controls (Fig. 5e).

COUP-TF2 also influenced coronary vessel proliferation. The relative increase in *Coup-tf2^{OE}* cells over developmental time was greater than for controls (Fig. 5f). Endothelial deletion of one copy of *Coup-tf2* resulted in decreased proliferation and expansion of coronary vessels (Fig. 5g and Extended Data Fig. 10e). As pre-artery specification was associated with decreased proliferation, these data suggest that COUP-TF2 may block arterial specification by activating cell cycle genes.

Next, we sought evidence that cell cycle exit enhances arterial specification, and that COUP-TF2 antagonizes this activity. First, cultured SV sprouts were treated with a cyclin-dependent kinase (CDK) inhibitor, which significantly increased artery differentiation (Extended Data Fig. 9l and m). Second, a CDK inhibitor was administered to *Cdh5CreER Coup-tf2*^{OE} mice dosed with tamoxifen early to assess whether the phenotype of small and absent coronary arteries could be alleviated (see phenotype in Fig. 4f). Inhibition of CDKs resulted in no significant difference between control and transgenic animals (Fig. 5h), demonstrating that the ability of COUP-TF2 to inhibit artery formation had been reversed.

Discussion

scRNA-seq can reveal developmental transitions at a much higher resolution than was previously possible 31-33. By combining scRNA-seq with in vivo localization and genetic manipulations, we show that a subset of endothelial cells within the immature coronary plexus crosses a transcriptional threshold to become pre-artery cells. Pre-artery specification is a critical step because blocking this process inhibited artery formation. Prior to pre-artery specification, SV-derived endothelial cells gradually decreased expression of venous genes while gradually increasing expression of arterial genes. These data suggest that fate switching during angiogenesis occurs in a progressive manner, and that individual plexus cells that reach a threshold towards full arterial differentiation form the mature coronary arteries.

Although COUP-TF2 is considered a master regulator of veins, precisely how it brings about venous fate and suppresses artery fate is still under investigation². Single-cell analysis revealed that COUP-TF2 did not push cells towards a venous fate or markedly suppress arterial genes. Instead, COUP-TF2 specifically blocked pre-artery specification, because *Coup-tf2*^{OE} induction before the pre-artery stage prevented mature artery development, whereas induction afterwards had little effect. Our data indicate that COUP-TF2 suppresses pre-artery specification by activating cell cycle genes. Recently, retinal artery differentiation has been shown to depend on cell cycle arrest triggered by blood flow, Notch activation, and CX37 (also known as GJA4)²⁷. Pre-artery specification was independent of flow, but may engage similar mechanisms. Future experiments should investigate whether this higher-resolution understanding of coronary artery differentiation during cardiac angiogenesis could aid the development of regenerative therapies.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0288-7

Received: 31 July 2017; Accepted: 29 May 2018; Published online 4 July 2018.

- Chen, X., Qin, J., Cheng, C.-M., Tsai, M.-J. & Tsai, S. Y. COUP-TFII is a major regulator of cell cycle and Notch signaling pathways. *Mol. Endocrinol.* 26, 1268–1277 (2012).
- Fish, J. E. & Wythe, J. D. The molecular regulation of arteriovenous specification and maintenance. Dev. Dyn. 244, 391–409 (2015).
- Isogai, S., Lawson, N. D., Torrealday, S., Horiguchi, M. & Weinstein, B. M. Angiogenic network formation in the developing vertebrate trunk. *Development* 130, 5281–5290 (2003).
- Red-Horse, K., Ueno, H., Weissman, I. L. & Krasnow, M. A. Coronary arteries form by developmental reprogramming of venous cells. *Nature* 464, 549–553 (2010).
- Xu, C. et al. Arteries are formed by vein-derived endothelial tip cells. Nat. Commun. 5, 5758 (2014).
- Kametani, Y., Chi, N. C., Stainier, D. Y. R. & Takada, S. Notch signaling regulates venous arterialization during zebrafish fin regeneration. *Genes Cells* 20, 427–438 (2015).
- Wu, B. et al. Endocardial cells form the coronary arteries by angiogenesis through myocardial-endocardial VEGF signaling. Cell 151, 1083–1096 (2012).
- Chen, H. I. et al. The sinus venosus contributes to coronary vasculature through VEGFC-stimulated angiogenesis. *Development* 141, 4500–4512 (2014).
- Volz, K. S. et al. Pericytes are progenitors for coronary artery smooth muscle. eLife 4. e10036 (2015).
- Ivins, S. et al. The CXCL12/CXCR4 axis plays a critical role in coronary artery development. Dev. Cell 33, 455–468 (2015).
- Sharma, B., Chang, A. & Red-Horse, K. Coronary artery development: progenitor cells and differentiation pathways. *Annu. Rev. Physiol.* 79, 1–19 (2017).
- Todorov, V. & Filzmoser, P. An object-oriented framework for robust multivariate analysis. J. Stat. Softw. 32, 1–47 (2009).
- Gokce, O. et al. Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. Cell Reports 16, 1126–1137 (2016).
- Rivera-Feliciano, J. et al. Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development* 133, 3607–3618 (2006).
- Zhang, H. et al. Endocardium minimally contributes to coronary endothelium in the embryonic ventricular free walls. Circ. Res. 118, 1880–1893 (2016).
- Chen, Q. et al. Endothelial cells are progenitors of cardiac pericytes and vascular smooth muscle cells. Nat. Commun. 7, 12422 (2016).
- Lin, C.-J., Lin, C.-Y., Chen, C.-H., Zhou, B. & Chang, C.-P. Partitioning the heart: mechanisms of cardiac septation and valve development. *Development* 139, 3277–3299 (2012).
- Adams, R. H. et al. Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* 13, 295–306 (1999).
- Sacilotto, N. et al. MEF2 transcription factors are key regulators of sprouting angiogenesis. Genes Dev. 30, 2297–2309 (2016).
- The Tabula Muris Consortium, Quake, S. R., Wyss-Coray, T. & Darmanis, S. Single-cell transcriptomic characterization of 20 organs and tissues from individual mice creates a Tabula Muris. Preprint at https://www.biorxiv.org/ content/early/2018/03/29/237446 (2018).
- Mack, J. J. et al. NOTCH1 is a mechanosensor in adult arteries. Nat. Commun. 8, 1620 (2017).
- Tian, X. et al. Vessel formation. De novo formation of a distinct coronary vascular population in neonatal heart. Science 345, 90–94 (2014).
- Chen, H. I. et al. VEGF-C and aortic cardiomyocytes guide coronary artery stem development. J. Clin. Invest. 124, 4899–4914 (2014).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci.* USA 102, 15545–15550 (2005).
- Akimoto, S., Mitsumata, M., Sasaguri, T. & Yoshida, Y. Laminar shear stress inhibits vascular endothelial cell proliferation by inducing cyclin-dependent kinase inhibitor p21(Sdi1/Cip1/Waf1). Circ. Res. 86, 185–190 (2000).
- Lin, K. et al. Molecular mechanism of endothelial growth arrest by laminar shear stress. Proc. Natl Acad. Sci. USA 97, 9385–9389 (2000).
- Fang, J. S. et al. Shear-induced Notch-Cx37-p27 axis arrests endothelial cell cycle to enable arterial specification. *Nat. Commun.* 8, 2149 (2017).
- You, L.-R. et al. Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. Nature 435, 98–104 (2005).
- Qin, J. et al. COUP-TFII inhibits TGF-β-induced growth barrier to promote prostate tumorigenesis. Nature 493, 236–240 (2013).
- Grieskamp, T., Rudat, C., Lüdtke, T. H.-W., Norden, J. & Kispert, A. Notch signaling regulates smooth muscle differentiation of epicardium-derived cells. *Circ. Res.* 108, 813–823 (2011).
- Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175–188 (2016).
- Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509, 371–375 (2014).



33. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).

Acknowledgements We thank S. Tsai, M.-J. Tsai, S. Evans, B. Zhou, T. Quertermous and L. Iruela-Arispe for mice; M. Miyanishi for assistance with fluorescence-activated cell sorting; R. Morganti and G. Gulati for assistance with scRNA-seq; L. O'Brien, D. Bergmann, and V. Greco for manuscript comments; and J. Ban for technical assistance. T.S. is supported by the NIGMS of the National Institutes of Health (T32GM007276). K.R.-H. is supported by the NIH/NHLBL (R01-HL128503) and the New York Stem Cell Foundation (NYSCF-Robertson Investigator). T.T.D. is supported by the NIH/NHLBL T32 (HL098049) and an AHA Postdoctoral Fellowship. E.C.B. is supported by the NIH (R01-GM037734, R01-Al130471) and the Department of Veterans Affairs.

Reviewer information *Nature* thanks R. Adams, C. Marr and A. Siekmann for their contribution to the peer review of this work.

Author contributions T.S., R.S., G.S. and K.R.-H. conceived the study. T.S. and R.S. captured cells and performed scRNA-seq. G.S. performed scRNA-seq computation. T.S., G.S. and K.R.-H. performed scRNA-seq analysis. G.D.

performed CXCR7–GFP and FBLN2/ADM in situ hybridization. S.D. performed P2/P6 postnatal analysis. S.R. performed EdU experiments. A.H.C. performed E14.5/E17.5 lineage quantification. A.P. performed the *Isl*1 experiment. B.R. performed GSEA. T.T.D. and W.A.R. provided *Coup-tt*2 flox mice. K.E.Q. and K.M.C. provided CXCR7–GFP mice. L.M. provided *Cx40CreER* mice. S.W. and G.L. provided adult scRNA-seq. T.S., G.S., and K.R.-H. prepared the manuscript. T.S. performed most wet lab experiments. R.S., E.C.B., I.W., S.Q., and K.R.-H. provided resources.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0288-7.

 $\begin{tabular}{ll} \textbf{Supplementary information} is available for this paper at https://doi.org/10.1038/s41586-018-0288-7. \end{tabular}$

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.R. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. All mice were used in compliance with Stanford University IACUC regulations. The following mouse strains were used: wild type (CD1, Charles River Laboratories, Strain Code #022), ApjCreER⁸, Rosa^{Coup-tj2OE29}, Rosa^{nTmG} Cre reporter (The Jackson Laboratory, Gt(ROSA)26Sor^{tm4(ACTB-tdTomato,-EGFP)Luo}/J, Stock #007576), Rosa^{NICD} (The Jackson Laboratory, Gt(ROSA)26Sor^{tm1(Notch1)Dam}/J, Stock #008159), Rosa^{tdTomato} Cre reporter (The Jackson Laboratory, B6.Cg-Gt(ROSA)26Sor^{tm9(CAG-tdTomato)Hze}/J, Stock #007909), Isl1^{MerCreMer34}, Cdh5CreER³⁵, Cx40Creer³⁶, Nfatc1^{Cre7}, Rosa^{Confetti} (The Jackson Laboratory, Gt(ROSA)26Sor^{tm1(CAG-Brainbow2.1)Cle}/J, Stock #013731), Coup-tf2 flox (Mutant Mouse Regional Resource Center, B6;129S7Coup-tf2^{tm2Tsa}/Mmmh, Stock #032805MU). Apln-lacZ³⁷, CXCR7-GFP (The Jackson Laboratory, C57BL/6-Ackr3^{tm1Litt}/J, Stock #008591), CXCL12-DsRed (The Jackson Laboratory, Cxcl12^{tm2.1Sim}/J, Stock #022458), VE-Cadherin-CreER³⁸. All mice were maintained on a mixed background.

Timed pregnancies were determined by defining the day on which a plug was found as E0.5. For Cre inductions, tamoxifen (Sigma-Aldrich, T5648) was dissolved in corn oil at a concentration of 20 mg/ml and was injected into the peritoneal cavities of pregnant dams. For cell cycle inhibition, 0.4 mg dinaciclib was dissolved in 2.6% DMSO (in PBS) and was injected into the peritoneal cavities of pregnant dams. Dosing and dissection schedules for individual experiments were: (1) E12.5 single-cell RNA sequencing: tamoxifen on E9.5 and E10.5, dissection on E12.5. (2) E14.5 single-cell RNA sequencing: tamoxifen at E11.5 and E12.5, dissection at E14.5. (3) ApjCreER Coup-tf2^{OE} experiments: tamoxifen at E9.5 and E10.5, dissected at E14.5 or E15.5 for coronary contribution quantification. Same dosing schedule, but dissected at E11.5 and E14.5 for recombination rate experiment (E11.5 only) and expansion experiment. Same dosing schedule, but dissected at E11.5, E12.5, or E13.5, was used for ventricular coverage visualization; tamoxifen at E11.5 and E12.5, dissected at E15.5 for capillary visualization in Extended Data Fig. 9. (4) Cx40Creer Coup-tf2^{OE} experiments: tamoxifen at E11.5 and E12.5 or E13.5, dissected at E15.5; for Extended Data Fig. 9i: tamoxifen at E11.5, dissected at E15.5. (5) Cdh5CreER Coup-tf2OE before pre-artery: tamoxifen at E11.5 and E13.5, dissected at E15.5. (6) Cdh5CreER Coup-tf2OE after pre-artery: tamoxifen at E13.5 and E15.5, dissected at E16.5. (7) Cdh5CreER Coup-tf2^{OE} dinaciclib experiment: tamoxifen at E11.5 and E13.5, dinaciclib at E12.5, dissected at E15.5. (8) Cx40Creer Rosa^{confetti}: tamoxifen at E12.5, dissected at E15.5. (9) ApjCreER Coup-tf2^{OE} NICD experiment: tamoxifen at E11.5 and E12.5, dissected at E15.5. (10) Cx40Creer RosatdTomato lineage tracing: tamoxifen at E11.5, dissected at E12.5, P7 or P8; tamoxifen at E10.5, dissected at E15.5; tamoxifen at E16.5, dissected at P8. (11) Cdh5CreER Coup-tf2 flox dosage: tamoxifen at E10.5, dissected at E12.5; tamoxifen at E11.5, dissected at E13.5 or E14.5. (12) ApjCreER lineage tracing in right or left coronary artery: tamoxifen at E9.5 and E10.5, dissected at E14.5 and E15.5. (13) Pre-artery cells/Slc45a4 in ApjCreER lineage vessels: tamoxifen at E9.5 and E10.5, dissected at E13.5. (14) Additional Cx40Creer and ApjCreER lineage-tracing experiments: see Extended Data Fig. 7. (15) VE-Cadherin-CreER Coup-tf2^{OE}: tamoxifen at E15.5 and E16.5, dissected at E17.5.

For additional *Cx40Creer Rosa*^{tdTomato} embryonic lineage-tracing experiment, pregnant dams were dosed via oral gavage with 1 mg 4-OH tamoxifen (Sigma-Aldrich H6278) at E11.5 and dissected at E12.5 (Extended Data Fig. 6f) or E15.5 (Fig. 2).

For postnatal lineage tracing at P2 and P6, tamoxifen was injected into the peritoneal cavity of the mother when the neonates were at P2 so that tamoxifen could be passed from the mother to the neonates through milk.

No statistical methods were used to predetermine sample size. For in vitro experiments, cultures were randomly chosen for different treatments and experiments were performed multiple times. Randomization was not relevant to our mouse experiments because genotypes/groups were determined by mouse genetics. Blinding was used in scRNA-seq and mouse experiments, except for lineage tracing, EdU experiments, Coup- $TF2^{OE}$ cell quantification and NICD quantification, where blinding was not possible because cells positive for certain markers (MYC tag, GFP, tdTomato, EdU) revealed the identities of the samples.

Cell isolation for scRNA-seq. E12.5 scRNA-seq. SV-derived cells were captured by fluorescence-activated cell sorting (FACS) of ApjCreER lineage-labelled cells (Cre expressed in SV). An experiment was performed once in which male ApjCreER Rosa^{mTimG} mice were crossed to CD1 females, who were dosed with tamoxifen at E9.5 and E10.5. Embryos were removed and placed into cold, sterile PBS at E12.5. The SVs of each of 27 GFP-positive hearts were microdissected away from the ventricles and pooled into a 300-μl mix consisting of 500 U/ml collagenase IV (Worthington #LS004186), 1.2 U/ml dispase (Worthington #LS02100), 32 U/ml DNase I (Worthington #LS002007), and sterile DPBS with Mg²⁺ and Ca²⁺. The ventricles of the 27 hearts were minced with forceps and pooled together in another 300 μl of the aforementioned mix. The pooled SVs and ventricles were then incubated at 37 °C, and gently resuspended every 7 min. After the incubation, 60 μl cold FBS followed by 1,200 μl cold sterile PBS were added and mixed into each tube. The samples were then filtered through a 70-μm cell strainer; the filter and the source

tube were washed with a total of 1,200 μl sterile PBS. Cells were then centrifuged at 400g at 4 °C for 5 min. Each cell pellet was then gently resuspended in 600 μl 3% FBS (in sterile PBS). Cells were centrifuged again at 400g at 4 °C for 5 min. Each pellet was then gently resuspended in 2,000 μl 3% FBS and 32 U/ml DNase I in sterile PBS. Cells were kept on ice until they were used for FACS.

DAPI (1.1 $\mu M)$ was added to the cells immediately before FACS. Single cells with a low DAPI signal, moderate PE-Texas Red signal and the highest Alexa-Fluor 488 signal were sorted using Aria II SORP (BD Biosciences). Each cell was sorted into a separate well of a 96-well plate containing 4 μl lysis buffer. Cells were spun down after sorting and stored at $-80\,^{\circ}\text{C}$ until cDNA synthesis. A total of 480 SV cells and 480 ventricular cells were sorted and processed for cDNA synthesis. Cells were analysed on the AATI 96-capillary fragment analyser, and a total of 915 cells that had sufficient cDNA concentration were barcoded and pooled for sequencing.

E14.5 scRNA-seq. The experiment was performed once following the same procedure as for E12.5 above unless otherwise noted here.

One thousand, one hundred and fifty-two FACS-captured coronary cells lineage-labelled with ApjCreER were collected from E14.5 hearts (SV cells were excluded and the later time point used to ensure sufficient numbers of $Coup-tf2^{OE}$ cells). To isolate $Coup-tf2^{OE}$ cells, male ApjCreER $Coup-tf2^{OE}$ mice were crossed to $Rosa^{mTmG}$ females who were dosed with tamoxifen at E11.5 and E12.5 and the embryos removed at E14.5. A total of 16 GFP-positive embryos from four litters were dissected for cell isolation and FACS. To isolate wild-type cells, male ApjCreER $Rosa^{mTmG}$ mice were crossed to CD1 females. Pregnant dams were dosed with tamoxifen at E11.5 and E12.5 and embryos removed at E14.5. A total of 12 GFP-positive embryos from three litters were sorted out and further dissected. For both the wild-type and the $Coup-tf2^{OE}$ samples, a few GFP-negative embryos were processed for dissection and cell isolation in the exact same manner to serve as a negative control for the GFP signal during FACS.

Cells with the highest Alexa-Fluor 488 signal, low DAPI signal, and low PE-Texas Red signal were sorted into lysis buffer. For Coup- $tf2^{OE}$, a total of 861 cells were sorted and processed for cDNA synthesis. For wild-type, a total of 608 cells were sorted and processed for cDNA synthesis. Of these, 1,152 passed cDNA fragment quality control (concentration >0.05 ng/ μ l) and were sequenced. Of those, 1,126 passed QC threshold (>1,000 genes, 10^5 mm10-aligned reads). In Coup- $tf2^{OE}$ embryos, 326 cells expressed the FLAG-Myc transgene and were compared to the 423 control cells that passed QC.

cDNA synthesis and library preparation for scRNA-seq. We used Smart-seq2 to perform scRNA-seq 39 . Poly-A mRNA in the cell lysate was converted to cDNA and amplified as described 39 . Amplified cDNA in each well was quantified using a high-throughput fragment analyser (Advanced Analytical). After quantification, cDNA from each well was normalized to the desired concentration range (0.05–0.16 ng/µl) by dilution, consolidated into a 384-well plate, and subsequently used for library preparation (Nextera XT kit; Illumina) using a semiautomated pipeline as described 40,41 . The distinct libraries resulting from each well were pooled, cleaned-up and size-selected using precisely $0.6 \times$ to $0.7 \times$ volumes of Agencourt AMPure XP beads (Beckman Coulter), as recommended by the Nextera XT protocol (Illumina). A high-sensitivity Bioanalyzer (Agilent) run was used to assess fragment distribution and concentrations of different fragments within the library pool. It is important to note that after pooling the libraries and before sequencing there is no PCR step in our protocol. Pooled libraries were sequenced on NextSeq 500 (Illumina).

To identify the broad cell subtypes present, in situ hybridization data on 52 genes from the Euroexpress 42 database were compared to expression levels in an rPCA plot of all cell in the data set, excluding erythrocytes (Extended Data Fig. 1j).

Cell subtypes in the *ApjCreER*-labelled populations were manually defined using gene expression patterns in manually selected PC plots derived from multiple iterative rounds of rPCA (iRPCA). There were two overall goals of iRPCA. The

first was to fully describe the cellular subtypes within an scRNA-seq data set while minimizing over-clustering of homogenous populations or continua, clustering based on cell cycle phase or technical artefacts/cell quality, and under-clustering of small subpopulations. The second goal was to preserve continuity or discreteness between subpopulations.

Our pipeline differed from standard pipelines in several ways. First, we used rPCA (rrcov::PcaHubert) in lieu of standard PCA. Second, we replaced default PC scores by those calculated by the sum of top 60 genes: PC.score = PC.pos – PC.neg (Extended Data Fig. 1b, c). These two parameters were used because they provided more clearly defined separations among cells with unique gene expression patterns (see Extended Data Fig. 1a, b and additional description in main text). Finally, we made frequent use of PC pos/neg biplots, which we defined by:

$$PC.pos = \sum_{i=1}^{30} \frac{g_{i,p}}{\max g_{i,p}}$$

PC.neg =
$$\sum_{i=1}^{30} \frac{g_{i,n}}{\max g_{i,n}}$$

Where $g_{i,p}$ are the top 30 genes by positive loading to the PC and $g_{i,n}$ by negative loading. These were used to identify and exclude cell cycle-associated PCs (described below in Identifying cell cycle-regulated genes) (Extended Data Fig. 1d) and to inspect for cell doublets (expected to have nearly equal levels, on a log scale, of the top markers for two distinct subpopulations; we did not see any in our data set, possibly owing to strict FACS gating on FSC-W and SSC-W and the large spacing of wells on standard 96-well plates) (Extended Data Fig. 1e).

Cell subtype clusters were assigned through the following process. After removing a small number of erythrocytes, all cells in the data set were used to calculate 15 PCs where the input was all genes minus those in our cell cycle category (see Identifying cell cycle-regulated genes) and the output was PC plots based on the sum of top 60 genes. Among the resulting 15 PC plots, one was manually chosen for further analysis based on the following criteria: 1. cells were well separated among the PC axes; 2. expression patterns of the top 60 genes revealed distinct populations or clusters; and 3. the PC was not highly correlated with cell cycle genes (see Identifying cell cycle-regulated genes) or number of genes detected (that is, technical artefact). Distinct cell populations within the selected PC were manually identified by their separation from other cells within the plots and strong correlation with distinct gene expression patterns. One (or more) distinct cell population was then removed, and another iteration was performed to calculate another set of PCs containing the decreased number of cells. Each of these subsequent iterations similarly involved, first, a PC calculation (10-15 PCs depending on step), then, a manual selection of one PC plot based on the above-described criteria, and, finally, within that selected PC the manual identification or removal of cell subpopulations based on the above-described criteria. These iterations ended when the calculated PCs revealed a single continuum that was arranged in a linear progression on the PC plots, which indicated the presence of only two groups of cells: one with high expression of one set of markers and the other with high expression of a second set of markers (Extended Data Fig. 1k). These last continua were separated into two groups, which comprised the final clusters. In this way, a single continuum was not overclustered into more than two groups.

Included in the custom R scripts are the exact steps by which we obtained all the reported clusters in the E12.5 data. In the first two rounds, rPCA (rrcov::Pca-Hubert, k = 15) was run using all genes expressed in >1 cell, filtered by removing ribosomal proteins by grep(Rp[ls]*), as well as Rn45s (also known as Rna45s5), Lars2, and Malat1. In all rounds after that, the list of 202 cell cycle genes described below was also removed from the gene list. In total, 20 rounds of iRPCA were performed to cluster cells into the 10 subpopulations in this work.

Pairwise discreteness test. To analyse the relationship between pairs of subpopulations of cells, the cells of the two subtypes are first projected onto a single axis of identity. For the purpose of the following description, these populations are referred to as A and B. To do this, cells are scored by their expression of the top differentially expressed genes between the two populations. Differential expression is calculated as log fold change, fractional difference (difference in fraction of A cells expressing minus the fraction of B cells expressing), and Wilcoxon P value; genes are filtered by fold change >0.2 (natural logarithm), fractional difference, >0.05, and $P < 10^{-3}$. The top n genes, sorted by fold change and fractional difference, are referred to as g_a (top n genes enriched in A) and g_b (top n genes enriched in B). The results do not vary much for n between 20 and 100 (Extended Data Fig. 1l, only low-confidence connections change). In this work, the gene list is pre-filtered by removing ribosomal genes (Rp[ls]*) and cell cycle genes (the list

of 202 cell cycle genes described below). Cells are then given a score x by their expression of these genes:

$$x_{\rm A} = \sum_{g \in g_{\rm A}} \frac{g}{\max g}$$

$$x_{\rm B} = \sum_{g \in g_{\rm B}} \frac{g}{\max g}$$

$$x = \frac{x_{A}}{\max x_{A}} - \frac{x_{B}}{\max x_{B}}$$

Where g is in \log_{10} counts per million (CPM) units and max is the maximum across all cells in the pair of subtypes. This scores cells along the axis of cell identity along A and B. The resulting distribution of cells along this axis is tested for discreteness, or a lack of intermediate cells, by the width of the largest gap between the two distributions. The statistic is calculated by the following procedure (Extended Data Fig. 1f):

- 1. The distribution is fitted to a Gaussian mixture model with two components, giving means μ_A and μ_B .
 - 2. Cells within the range (μ_A , μ_B) are identified as candidate intermediates.
- 3. The largest gap distance between candidate intermediate cells, $d_{\rm max}$, is identified.
- 4. The list of candidate intermediate cells is further restricted to the 10 cells on either side of $d_{\rm max}$, and their gap distances, excluding $d_{\rm max}$, are fit to an exponential with rate k, F(d;k). If there is a uniform distribution of intermediate cells along the continuum from A to B, the gap distances d_i follow an exponential distribution $P(d) \approx e^{-kd}$, where the mean gap distance E[d] = 1/k (equivalent to the mean time between events for a Poisson process occurring at rate k).
- 5. The discreteness statistic is calculated as $D = \log_{10} F(d_{\text{max}}; k)$.
- 6. Two populations are considered discrete if D < -6. In the PlotConnectogram function, distributions with -3 > D > -6 are connected by a semitransparent lines to indicate lower confidence in their continuity. In simulated data, this corresponded to 3–5 intermediate cells. Distributions with med(D) > -3 are connected by 100%-opacity lines to indicate high confidence in their continuity.

Estimating the number of intermediate cells. Second, the number of intermediate cells connecting the two pairs is estimated by maximum-likelihood fitting of a five-parameter distribution. This distribution was derived by considering two cell types with mean expression values μ_A , μ_B and a transitional population sampled evenly from the range of values $\mu_A < \mu < \mu_B$. The exact PDF that describes sampling from this distribution with Gaussian noise is:

$$P(x;\mu_{\mathrm{A}},\mu_{\mathrm{B}},\sigma) = f_{\mathrm{A}}N(x;\mu_{\mathrm{A}},\sigma) + f_{\mathrm{B}}N(x;\mu_{\mathrm{B}},\sigma) + f_{\mathrm{AB}}\int_{\mu_{\mathrm{A}}}^{\mu_{\mathrm{B}}}N(x;\mu,\sigma)d\mu \qquad (1)$$

where f_A is the fraction of cells in cell type A, f_B is the fraction of cells in cell type B, and f_{AB} is the fraction of cells along the A–B continuum. The integral in (1) is approximated by

$$\int_{\mu_{\rm A}}^{\mu_{\rm B}} N(X;\mu,\sigma) d\mu \approx \begin{cases} 0, & x < \mu_{\rm A} - 2.7\sigma \\ C \exp\left[\frac{(x - \mu_{\rm A})^3}{3\sigma^3}\right] + D, & \mu_{\rm A} - 2.7\sigma \le x < \mu_{\rm A} + 2.7\sigma \\ 1, & \mu_{\rm A} + 2.7\sigma \le x < \mu_{\rm B} - 2.7\sigma \end{cases} \tag{2}$$

$$F \exp\left[\frac{(x - \mu_{\rm B})^3}{3\sigma^3}\right] + G, \quad \mu_{\rm B} - 2.7\sigma \le x < \mu_{\rm B} + 2.7\sigma \\ 0, & x \ge \mu_{\rm B} + 2.7\sigma \end{cases}$$

Where C, D, F, G are calculated to make (2) a continuous function. This PDF is then fit to the distribution using Nelder-Mead optimization (stats::optim) with five iterations for different initial values of f_{AB} . The initial values for μ_A , μ_B , and σ are derived by fitting with a two-component Gaussian mixture model. f_{AB} determines the width of the lines connecting populations in our PlotConnectogram function. Simulation of population distributions for model validation. We optimized the cutoffs for the discreteness test using simulated data. The data was simulated by drawing from the five-parameter distribution described above under Estimating the number of intermediate cells, where f_{AB} ranged from 0 to 1 (Extended Data Fig. 1h). Using the simulations, we found -6 to be a good cutoff for calling cell types discrete—this cutoff is low so as to be sufficiently sensitive to a small number of intermediates (\sim 3 intermediate cells out of 150).

Identifying cell cycle-regulated genes. When mentioned in the main text, we filtered out a list of 202 cell cycle genes from the input to rPCA to reduce the contribution of cell cycle to heterogeneity. We defined this list by rPCA: cell cycle PCs were identified by high loadings of known cell cycle markers (for example, cyclins, Mki67, Top2a). Also, cell cycle has a unique pattern on PCi.pos versus PCi.neg biplots (described above): there is typically a large coordinated increase in genes upon entering cell cycle with little corresponding decrease in genes, and PCi.pos has low correlation to PCi.neg. The positive and negative loadings were therefore inspected separately for cell cycle genes. In this work, rPCA was performed on a highly cycling, relatively homogeneous subgroup of cells (later identified as SVc and CV) using all genes; we used the union of the top 60 genes by each of the following loadings: PC1-positive, PC2-negative, PC2-negative, PC4-positive, PC5-negative, and PC6-negative, which produced a list of 230 candidate cell cycle genes. We filtered this list for genes that had high loadings to other PCs, marked subpopulations of cells, and had no cell cycle annotation; these included arterial markers such as *Unc5b*. This produced the final list of 202 cell cycle genes. This list was not complete, but was sufficient to remove cell cycle heterogeneity from the top PCs. Defining the fetal SVc-CV plexus-arterial axis. We defined the SVc-CV plexusarterial axis (x) using the scores generated by PC2 and PC3 from RPCA on SVc, CV, and arterial cells (Fig. 3a, Extended Data Fig. 4) as below:

$$x = \begin{cases} PC2.score, PC3.score^3 - PC2.score - 0.4\\ \sqrt{PC2.score^2 + PC3.score^2}, PC3.score < -PC2.score - 0.4 \end{cases}$$

Figure 3a was coloured by the value of this axis.

Cell cycle scoring. G1/S and G2/M signatures were discovered in an unbiased manner as follows: coronary vessel plexus cells from wild-type E14.5 animals were analysed with rPCA using all detected genes. Many of the top 60 genes by loading to PC3.neg and PC2.neg were known G1/S markers, and, thus, the G1/S score of a cell was defined by the sum of the scaled expression of these genes. Many of the genes with high loadings to PC4.neg and PC5.neg were known G2/M markers, and the G2/M score was calculated by the sum of the scaled expression of these genes. Cells were scored as cycling if they were not in the bottom-left modes (high expression of at least one cell cycle signature).

Seurat clustering for comparison. To compare our clustering to Seurat, we ran Seurat with primarily default options. We filtered our list of 202 cell cycle genes as well as ribosomal proteins from the list of highly variable genes (y.cutoff = 0.5) and ran PCA with 20 scores calculated. Based on the PC elbow plot, we selected the first 10 PCs to be used for clustering. We excluded PC6 for high loading of cell cycle genes (since our list of 202 genes was not exhaustive), and clustered using FindClusters with resolution 2. We also calculated t-SNE, and we used the t-SNE mediods of cell clusters to place the vertices for our results from pairwise PCA (Extended Data Fig. 11).

Comparison to adult artery-vein continuum. We determined the similarity between our E12.5 endothelial cells and the mature artery-vein continuum as follows. We selected cells from the *Tabula Muris* data set with the Tissue label 'Heart' and annotation label '1'. We ran PCA on the most variable genes (y.cutoff=0.35) with the Seurat package. PC2 and PC3 separated cells into three populations along a single continuum, and we projected cells onto a single axis $x = \begin{cases} PC2, PC2 < 0 \\ \sqrt{PC2^2 + PC3^2}, PC2 > 0 \end{cases}$. Known arterial genes such as *Cx40*, *Cx37*, and

Unc5b were negatively correlated, and known capillary/venous markers such as *Apj* and *Nrp2* were positively correlated to the axis, so we considered it to be the artery–vein continuum (AVc). We then calculated the similarity of each fetal cell to each adult cell. To do this, we used as input the union of the top 300 genes correlated to the adult and fetal AVc, smoothed by LOESS regression over the AVc defined above. We calculated the Pearson correlation similarity using these features, and mapped each fetal cell to the adult cell to which it was most similar by this metric.

Immunohistochemistry and Imaging. For whole-mount embryonic hearts. All embryos were fixed in 4% PFA at $4\,^{\circ}\text{C}$ with shaking and washed twice (10 min each wash) with PBS at room temperature with shaking before dissection for whole-mount immunostaining.

Intact embryonic hearts were washed in PBT (PBS with 0.5% Triton-X 100) at room temperature for one hour before incubation with primary antibodies. Primary antibodies were dissolved in either 5% goat serum or 5% donkey serum in PBT. Hearts were incubated in the solution with primary antibodies with shaking overnight at 4 °C. Hearts were then washed with PBT for six to nine hours with shaking at room temperature, and the wash was changed every hour. Hearts were then stained with secondary antibodies with the same conditions and procedure as for primary antibodies. After washing off the secondary antibodies, hearts were then left in enough PBT to cover them. Two drops of Vectashield (Vector Labs, H1000) were added and mixed with the PBT for each heart, and the hearts were

stored at $-20\,^{\circ}$ C for the long term. Imaging was done with Zeiss LSM-700 ($10\times$ or $20\times$ objective lens) with Zen 2010 software (Zeiss).

For whole-mount postnatal hearts. Hearts were fixed in 4% PFA for 1 h at 4°C with shaking and washed twice (15 min each wash) with PBS at 4°C with shaking before dissection for whole-mount immunostaining. In the primary antibodies (diluted in PBT), hearts were shaken at room temperature for 6 h and overnight at 4°C. To wash the primary antibodies, hearts were shaken in PBT at room temperature for 10 h and overnight at 4°C. Hearts were washed in 50 ml PBT and the wash was changed every 2 h while shaking at room temperature. Hearts were then placed in secondary antibodies (diluted in PBT) at room temperature with shaking for 6 h and overnight with shaking at 4°C. Hearts were then washed in 50 ml PBT for 8 h (wash changed every 2 h) and overnight at 4°C. The washing was repeated for six more days. Prior to imaging, Vectashield (Vector Labs, H1000) was added to hearts in clean tubes, and hearts were equilibrated at room temperature for 40 min. Imaging was done with Zeiss LSM-700 (10× or 20× objective lens) with Zen 2010 software (Zeiss).

Primary and secondary antibodies. The following primary antibodies were used at the indicated concentrations: MYC-Tag for COUP-TF2^{OE} (Cell Signaling Technology, Inc., 2278S, 1:300), VE-Cadherin (BD Pharmingen, 550548, 1:125), VEGFR2 (R&D Systems, AF644, 1:125), CX40 (Alpha Diagnostic International, CX40A, 1:300), ERG (Abcam, ab92513, 1:500), CXCR4 (BD Pharmingen, 551852, 1:125), GFP (Abcam, ab13970, 1:500), VWF (Abcam, ab6994, 1:500), CLDN11 (Abcam, ab53041, 1:1,000), SOX17 (R&D Systems, AF1924, 1:500), anti-actin α-smooth muscle-FITC (Sigma, F3777, 1:200), VEGFR3 (R&D Systems, AF743, 1:125), DACH1 (Proteintech, 10914-1-AP, 1:500), JAG1 (R&D Systems, AF599, 1:125).

All secondary antibodies were Alexa Fluor conjugates (488, 555, 633, 635, 594, 647, Life Technologies, 1:125 or 1:250). DAPI (1 mg/ml) was used at 1:500.

In situ hybridization. To identify the broad cell subtypes in the E12.5 single cell data set, expression levels in rPCA plots of 52 genes were compared to in situ hybridization data from the Euroexpress⁴² and Allen Brain Atlas databases (stages ranged from E11.5 to E15.5). Expression patterns from E14.5 Euroexpress data are shown in Extended Data Fig. 1i.

For Adm and Fbln2, in situ hybridization on paraffin sections were performed twice as described previously⁴³. Antisense Adm and Fbln2 probes were labelled with digoxigenin (DIG)-UTP using the Roche DIG RNA labelling System according to the manufacturer's guidelines.

For *Slc45a4*, whole hearts were fixed and in situ hybridization performed according to protocol from Additional File 2 of ref. ⁴⁴. Probes were *Cdh5* (Advanced Cell Diagnostics 312531-C2), *Cx40* (Advanced Cell Diagnostics 518041), and *Slc45a4* (Advanced Cell Diagnostics 522131-C3). Reagents are RNAscope Protease III & IV Reagents (Advanced Cell Diagnostics 322340) and RNAscope Fluorescent Multiplex Detection Reagents (Advanced Cell Diagnostics 320851). About 12 embryonic hearts were dissected in a sterile and RNase-free environment into a 1.5-ml tube and fixed in 1 ml 4% PFA for 1 h at room temperature. Three fixed hearts were processed in the same tube with 100 μ l of the probes master mix. The experiment was performed three times, once each for E13.5 (n=3), E14.5 (n=2), and E15.5 (n=3).

SV-atria explant experiment. The experiment was performed three times. In total, 71 embryos were dissected at E12.5. The SV and atria of each embryo were dissected on sterile PBS and gently dropped onto a cell culture insert (EMD Millipore PI8P01250) coated with Matrigel (BD Biosciences) inside a well of a 24-well plate. Two to five explants were cultured onto each insert. Immediately after the explants were dropped onto the insert, 200 μl EGM2-MV medium was added into the space between the insert and the well. The SVs were allowed to attach onto the Matrigel at 37 °C for 2–6 h before another 200 μl EGM2-MV medium was added to the space between the insert and the well. The explants were cultured at 37 °C for approximately 72 h before either flavopiridol or DMSO was added: 900 μl of 40 nM flavopiridol (dissolved in 0.1% DMSO in EGM2-MV) or 0.1% DMSO in EGM2-MV (drug vehicle control) was added to each insert. After addition of either flavopiridol or DMSO, explants were incubated at 37 °C for approximately 48 h before they were fixed and stained.

Each cell culture insert was fixed in 1,000 μ l 4% PFA for 2 h at 4°C without shaking. Then, each insert was washed with 1,000 μ l PBS three times at room temperature. Five hundred microlitres of primary antibodies (diluted in 0.5% PBT) were added onto each insert and inserts were incubated at room temperature with shaking for 4–6 h. The inserts were subsequently washed with PBS at room temperature with shaking for 2 h. Five hundred microlitres of secondary antibodies (diluted in 0.5% PBT) were added onto each insert and inserts were incubated at 4°C for about 16 h. The inserts were then washed with PBS three times at room temperature with shaking for 2 h. The membrane containing the SVs was then excised from the insert and mounted onto a drop of Vectashield on a slide and stored at -20°C. Imaging was done using a Zeiss LSM-700 ($10 \times$ or $20 \times$ objective lens) with Zen 2010 software (Zeiss).

Acquisition and processing of images. All images were acquired with Zen 2010 software (Zeiss). Images were prepared using Photoshop CS6 (Adobe). Any changes to brightness and contrast were applied equally across the entire image. In vivo EdU Assay. To measure in vivo proliferation rate, $50\,\mu\text{g/g}$ body weight of EdU was injected into pregnant mice intraperitoneally 2–3 h before embryo collection. EdU-positive cells were detected using a Click-iT EdU kit (Invitrogen, C10338) according to the manufacturer's instructions. In brief, Click-iT reaction cocktails were incubated for 30 min after the secondary antibody incubation of the immunostaining protocol.

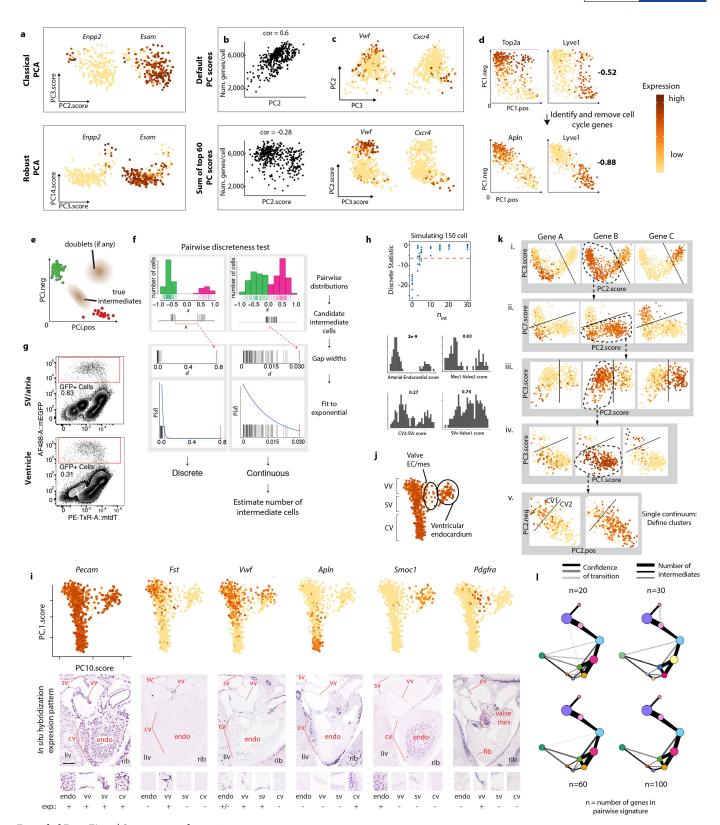
Quantification and statistical analysis of confocal images. See Supplementary Methods for details.

Code availability. The custom R scripts used to analyse the scRNA-seq data are publicly available on GitHub (https://github.com/gmstanle/coronary-progenitor-scRNAseq).

Data availability. Raw scRNA-seq data are available at https://github.com/gmstanle/coronary-progenitor-scRNAseq. Figures associated with the raw data are Figs. 1, 3, 4, and Extended Data Figs. 1–8. There is no restriction on data availability. Source Data for Figs. 2, 4, 5, and Extended Data Figs. 6–9 are provided with the paper.

34. Sun, Y. et al. Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Dev. Biol.* **304**, 286–296 (2007).

- 35. Wang, Y. et al. Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
- Miquerol, L. et al. Endothelial plasticity drives arterial remodeling within the endocardium after myocardial infarction. Circ. Res. 116, 1765–1771 (2015).
- Sheikh, A. Y. et al. In vivo genetic profiling and cellular localization of apelin
 reveals a hypoxia-sensitive, endothelial-centered pathway activated in ischemic
 heart failure. Am. J. Physiol. Heart Circ. Physiol. 294, H88–H98 (2008).
- Alva, J. A. et al. VE-Cadherin-Cre-recombinase transgenic mouse: a tool for lineage analysis and gene deletion in endothelial cells. *Dev. Dyn.* 235, 759–767 (2006).
- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protocols 9, 171–181 (2014).
- Koh, P. W. et al. An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. Sci. Data 3, 160109 (2016).
- Loh, K. M. et al. Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types. *Cell* **166**, 451–467 (2016).
- 42. Diez-Roux, G. et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011).
- D'Amato, G. et al. Sequential Notch activation regulates ventricular chamber development. Nat. Cell Biol. 18, 7–20 (2016).
- Gross-Thebing, T., Paksa, A. & Raz, E. Simultaneous high-resolution detection of multiple transcripts combined with localization of proteins in whole-mount embryos. *BMC Biol.* 12, 55 (2014).

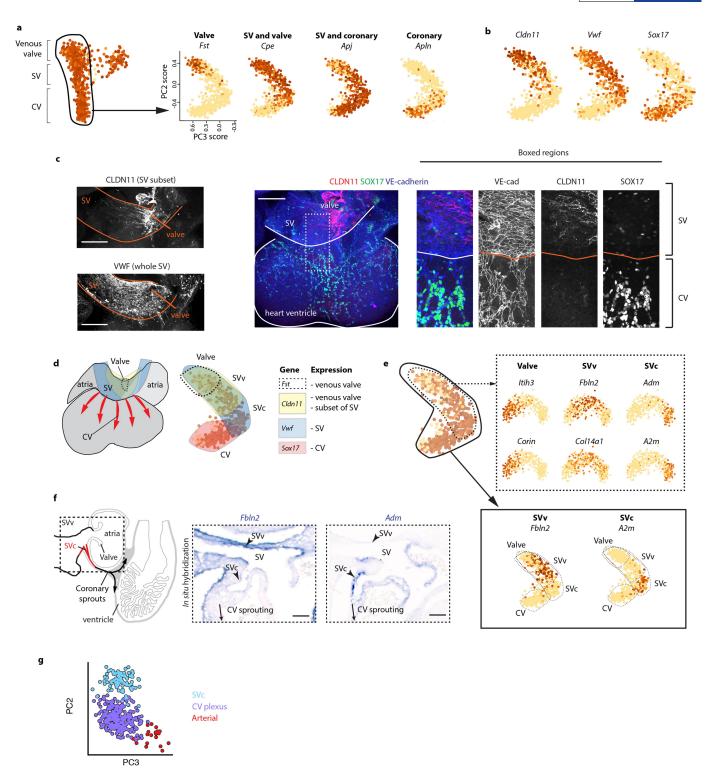


Extended Data Fig. 1 \mid See next page for caption.

RESEARCH ARTICLE

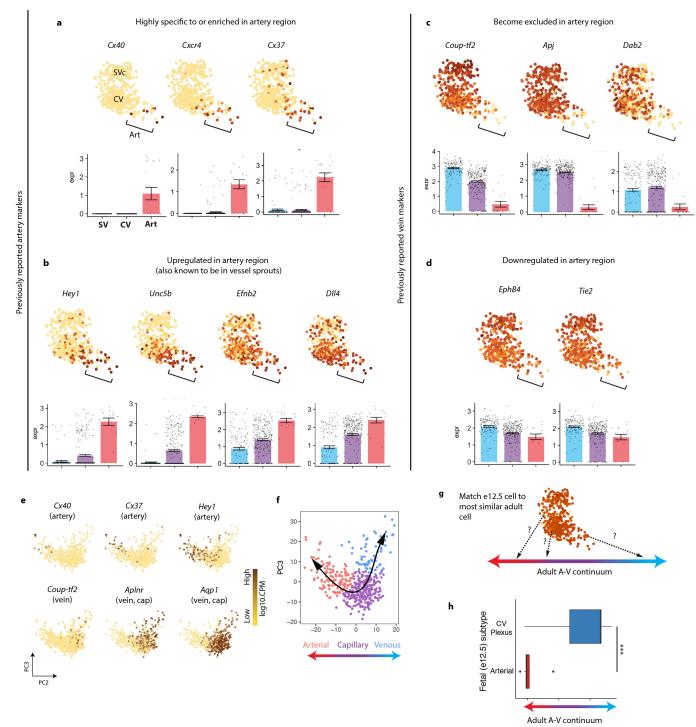
Extended Data Fig. 1 | Single cell analysis of ApjCreER lineage labelled cells. a, Comparison of rPCA and classical PCA at separation of subpopulations. PC scores were selected to best separate the Enpp+ Esam⁻ population. Cells are coloured by expression (log₁₀ CPM, scaled to maximum per gene). n = 352 cells. **b**, Comparison of default and sum-of-60 modified PC scores. PC2 is the default PC score from rPCA; PC2.score is the modified sum-of-top-60 scores (expression is log_{10} CPM, scaled to maximum). Y-axis is the number of genes detected per cell (>1 count). n = 426 cells. **c**, Comparison of default and sum-of-top-60 scores. Scores were chosen that best separated the *Vwf*⁺ and *Cxcr4* populations n = 426 cells. **d**, Unique cell cycle signature on PC.pos/PC.neg biplots. PC1.pos (PC1.neg) is the sum of the top 30 genes by positive (negative) loading to PC1. Cells are coloured by expression. Lower panel is the same rPCA after removing the list of 202 cell cycle genes. Numbers in bold are the correlations between PC1.pos and PC1.neg. n = 674 cells. **e**, PC.pos/ PC.neg biplot showing theoretical location of doublets expressing high levels of both gene sets. f, Schematic of the pairwise discreteness test on a

discrete (left) and continuous (right) pair of subpopulations. g, FACS plots used to isolate GFP-positive cells (red box) from ApjCreER $Rosa^{mTmG}$ hearts at E12.5. **h**, Top, discreteness statistic generated by pairwise discreteness test as a function of number of intermediate cells (n_{int}) for simulated distributions. Bottom, pairwise distributions of cell clusters in the data set and the fraction of intermediate cells estimated by pairwise discreteness analysis. i, rPCA plots and their accompanying gene expression patterns in the embryonic heart as reported by Euroexpress. In situ hybridization images show whole hearts (top); insets of specific areas are in lower panels with relative expression levels indicated. Expression levels in rPCA plots range from 0 (yellow) to 4 (brown) in log₁₀CPM. Top, n = 843 cells. **j**, Summary of broadly defined cell populations as indicated by gene expression patterns. n = 843 cells. k, Example of manual clustering process. For i, n = 732 cells; ii, n = 531 cells; iii, n = 415 cells; iv, n = 284cells; v, n = 261 cells. l, Comparison of pairwise discreteness test results for different numbers of genes per cell type signature (n).



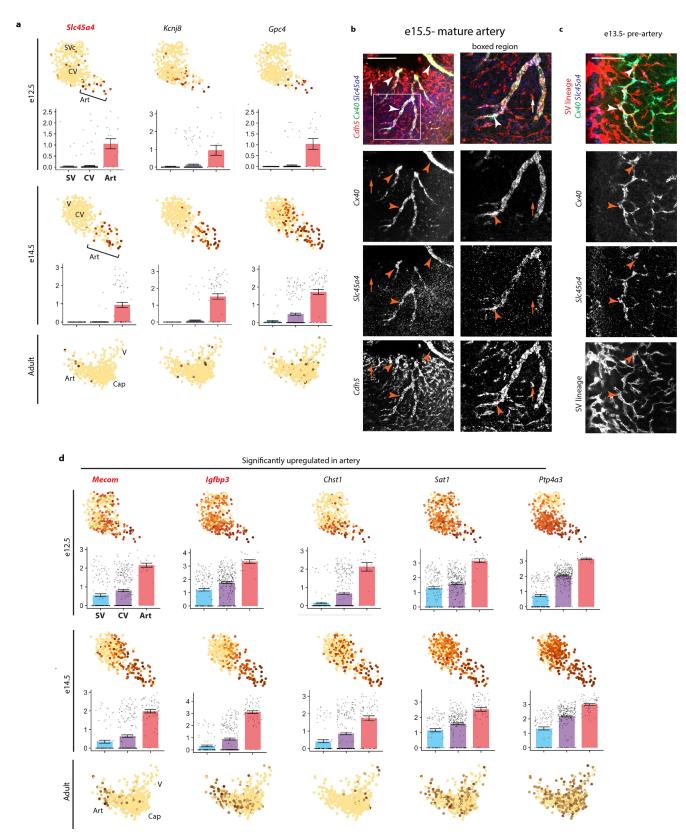
Extended Data Fig. 2 | Identification of a coronary progenitor niche within the SV. a, Gene expression patterns identify cell types in rPCA plots of the venous valve–SV–CV continuum. Expression levels are \log_{10} CPM and range from 0 (yellow) to 4 (brown) as indicated. Left, n=843 cells; right, n=732 cells. b, c, Expression patterns in rPCA plots (b) and whole-mount confocal immunofluorescence (c) of selected genes. For b, n=732 cells. d, Overlaying gene expression patterns suggests that the

SV has two distinct domains, the SVc (sinus venosus, coronary adjacent) and the SVv (sinus venosus, valve adjacent). **e**, rPCA on the valve–SVv–SVc continuum identified specific markers of the SVv and SVc. Solid box, n=732 cells. Dotted box, n=415 cells. **f**, In situ hybridization of SVv and SVc markers revealed complementary localization in vivo. **g**, Colour coding showing subpopulations that were used to calculate average expression levels. Scale bars, **b**, 200 μ m, **e**, 30 μ m.



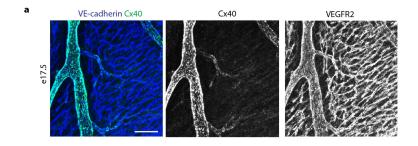
Extended Data Fig. 3 | Characterization of pre-artery cells. a–d, rPCA plots of the E12.5 SVc–CV continuum. Each dot is an individual cell, and gene expression levels are indicated by the colour spectrum as shown in Fig. 1d, which reflects \log_{10} CPM. a, Arterial genes highly enriched in the arterial areas of the plot. b, Arterial genes significantly upregulated in, but not specific to, the arterial area of the plot. c, Venous genes highly depleted in the arterial areas of the plot. d, Venous genes downregulated, but not depleted, in the arterial area of the plot. For a–d, Bonferroniadjusted P < 0.01; PCA plots, n = 415 cells. Centre and error bars are mean \pm s.e.m. of log CPM expression values. e, Genes expressed in adult

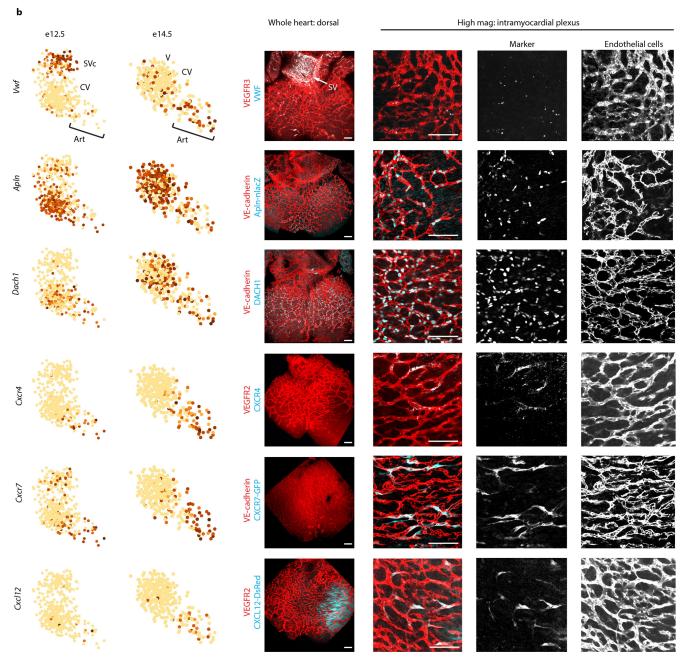
coronary artery cells. Data are from the *Tubula Muris* consortium. n=445 cells. **f**, Assignment of artery, capillary, and vein in adult coronary cells based on gene expression enrichment in **e**. n=445 cells. **g**, Schematic for comparing E12.5 coronary cells to those along the adult artery–capillary–vein continuum. **h**, Results of experiment schematized in **g**. The centre line correspond to the median; the upper and lower hinges correspond to the first and third quartile, respectively; the whiskers extend to the largest value or to $1.5 \times \text{IQR}$ (inter-quartile range, or distance between quartiles), whichever is smaller. Pre-artery cells: n=20 cells. CV: n=277 cells. $P=6.2 \times 10^{-13}$. Statistical test is two-tailed.



Extended Data Fig. 4 | Novel artery markers identified in scRNA-seq data. a, E12.5, E14.5, and adult coronary cell rPCA plots with genes highly enriched or specific to the arterial area during development. Each dot is an individual cell, and gene expression levels are indicated by the colour spectrum as shown in Fig. 1d, which reflects \log_{10} CPM. Genes in bold red are also enriched in adult artery cells. b, Fluorescence in situ hybridization (RNAscope) for Slc45a4, which is expressed (arrowheads) in vessels positive for the arterial marker Cx40, but not in Cx40-negative capillaries (arrows). c, Slc45a4 expression in pre-artery cells derived from the SV

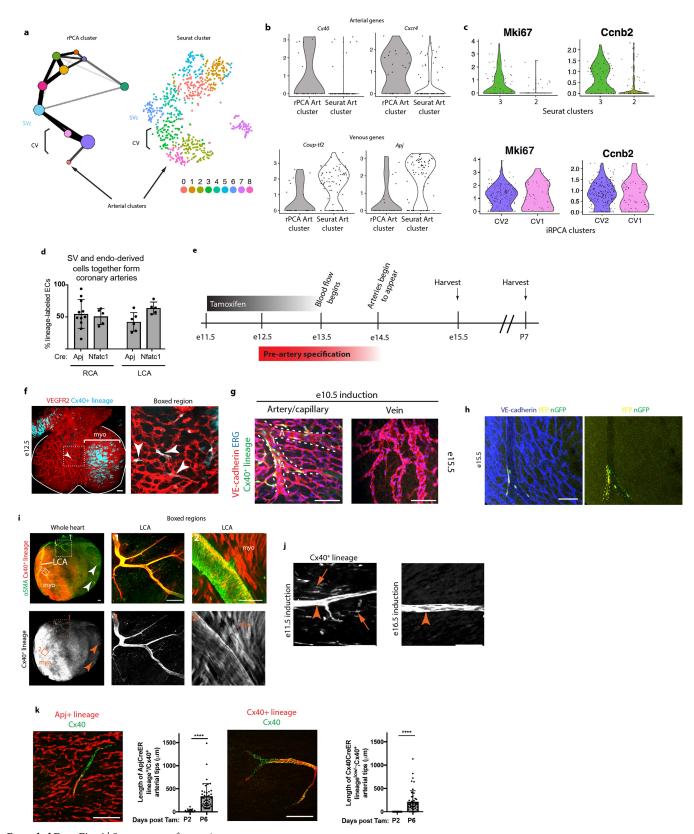
lineage (*ApjCreER* lineage-labelled; arrowheads). **d**, Genes enriched in, but not specific to, arterial cells at E12.5 and E14.5. Genes in bold red are arterial specific in both the developing and adult heart. In **a** and **d**: for PCA plots, n = 415 cells (top, E12.5); n = 347 cells (middle, E14.5); n = 445 cells (bottom, adult). For bar graphs E12.5, Art, n = 20 cells; CV, n = 277 cells; SV, n = 118 cells; E14.5, Art, n = 70 cells; CV, n = 454 cells; SV, n = 144 cells. Centre and error bars are mean \pm s.e.m. of log CPM expression values. Dots represent individual cells. Scale bars, $100 \,\mu\text{m}$.





Extended Data Fig. 5 | Additional whole mount immunofluorescence of marker genes. a, CX40 whole-mount immunohistochemistry in late gestation hearts (E17.5). CX40 is expressed only in cells lining large arteries and arterioles (overlapping blue and green signal). Low level, non-arterial signal is in myocardial cells. b, rPCA plots from E12.5 and E14.5 with accompanying whole-mount immunofluorescence in E13.5

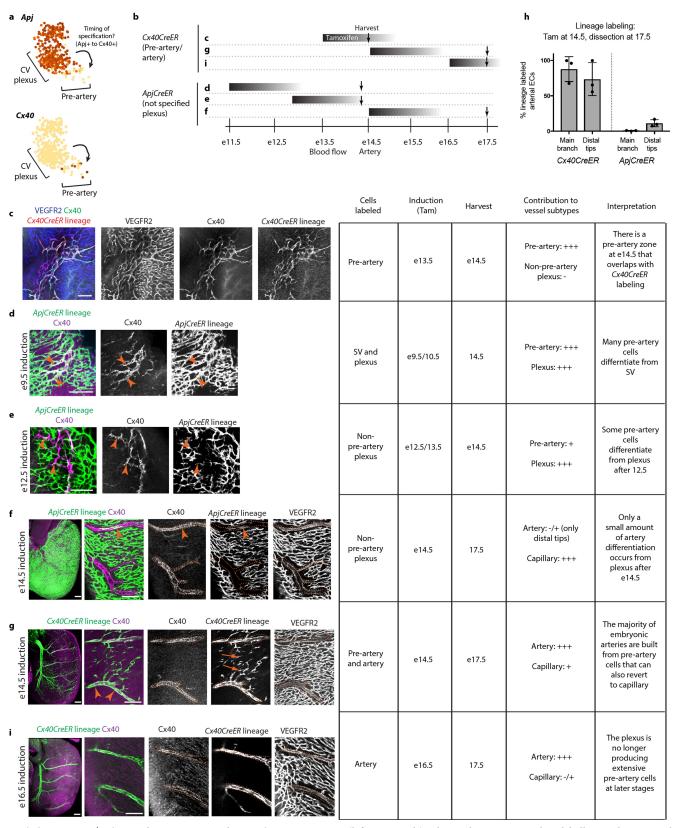
hearts. VWF is enriched in the SV; while APLN–nlacZ signal and DACH1 are present throughout the coronary plexus. CXCR4, ACKR3–GFP, and CXCL12–DsRed are enriched in the pre-artery and artery areas of rPCA plots and are interspersed within the intramyocardial coronary plexus. n=415 cells (left, E12.5); n=347 cells (right, E14.5). Scale bars, 100 μ m.



Extended Data Fig. 6 \mid See next page for caption.

Extended Data Fig. 6 | Clustering and additional lineage analysis of **pre-artery cells. a**, Clusters and relationships generated by rPCA and the pairwise discreteness test (left) and clusters generated by the Seurat pipeline (Louvain/SNN clustering, resolution = 2) (right). n = 757 cells. b, Violin plots show that arterial gene enrichment and venous gene deenrichment are better with manual, iterative clustering, suggesting that this method leads to more precise populations. c, Violin plots of cell cycle genes in the two CV plexus clusters generated by the indicated algorithms. Seurat clusters are more defined by cell cycle differences than iterative rPCA (iRPCA) clusters. b, c, Violin plots were made using Seurat VlnPlot. Each violin plot is one subtype and each dot corresponds to a cell. \mathbf{d} , Quantification of SV and endocardium contributions to coronary arteries. Error bars show s.d. ApjCreER RCA: n = 11 hearts. ApjCreERLCA: n = 6 hearts. Nfatc1Cre RCA: n = 5 hearts. Nfatc1Cre LCA: n = 5hearts. Centre, mean. e, Experimental design to lineage trace pre-artery cells. f, Lineage labelling in E12.5 Cx40CreER Rosatidtomato hearts induced with tamoxifen at E11.5. g, Arterial lineage labelling in hearts induced at E10.5. **h**, Example of clones in *Cx40CreER Rosa^{confetti}* heart at E15.5. Tamoxifen was administered at E12.5. Two groups of cells sharing the same fluorescent label (clones) are present: YFP-labelled (yellow circle)

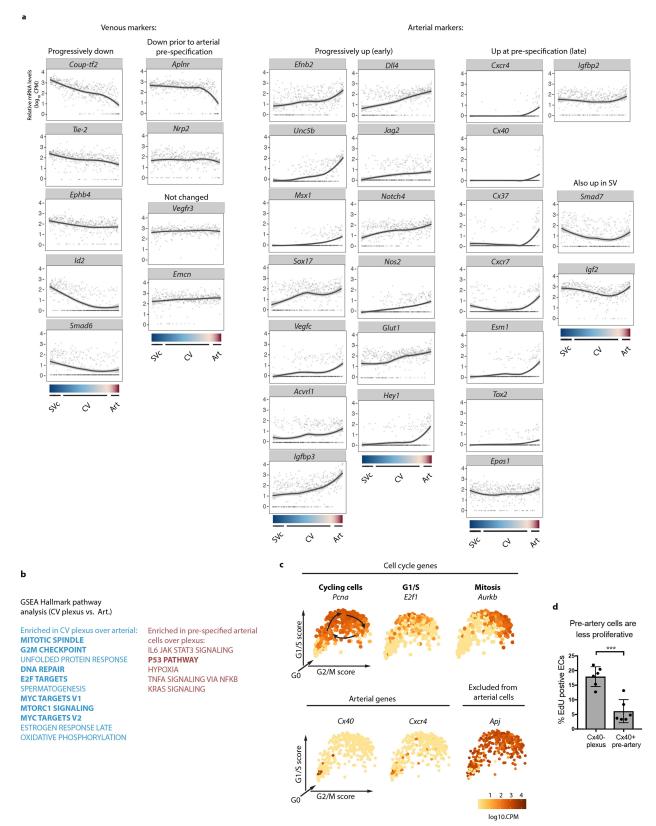
and nGFP labelled (green circle). Clone sizes are very small, consistent with low proliferation rates in pre-arterial cells. i, P8 heart lineage from Cx40CreER Rosa^{tdtomato} mice dosed with tamoxifen at E11.5. Heavy lineage labelling of the left coronary artery is shown (LCA). Arrowheads indicate branches of the right coronary artery. Myocardium (myo) of the left ventricle is also $Cx40^+$ at E11.5, and is also lineage labelled. **j**, Images from P8 Cx40CreER $Rosa^{tdtomato}$ hearts dosed with tamoxifen at E11.5 or E16.5. Only the E11.5 dosage results in capillary labelling (arrows) resulting from reversion of pre-artery cells that differentiate during the burst of pre-artery specification between E12.5 and E14.5. Arrowheads point to arterial lineage labelling. k, Postnatal lineage tracing in ApjCreER $Rosa^{tdtomato}$ or Cx40CreER $Rosa^{tdtomato}$ hearts where tamoxifen was injected at P2. Tips of arteries are lineage labelled with ApjCreER Rosatdtomato, but are depleted of Cx40CreER Rosatdtomato label, indicating that artery tips can extend by incorporating capillary cells that differentiate into arterial endothelial cells. Unpaired two-tailed *t*-test was used to calculate *P* values. For ApjCreER, n = 78 artery tips at P2, n = 41 artery tips at P6. P = 4.4608 \times 10⁻¹⁹. For *Cx40CreER*, n = 81 artery tips at P2, n = 49 artery tips at P6. $P = 1.61705 \times 10^{-15}$. Error bars show s.d. **** $P \le 0.0001$. Centre is mean. Scale bars, \mathbf{d} , \mathbf{e} , \mathbf{f} , \mathbf{j} , \mathbf{k} , 100 μ m; \mathbf{g} , 50 μ m.



Extended Data Fig. 7 | A burst of pre-artery specification between E12.5 and E14.5 specifies cells that build most of the embryonic left and right coronary arteries. a, rPCA plots of the SVc-CV continuum show that Apj and Cx40 mark cells before and after pre-artery specification, respectively. n=415 cells. b, Schematic of lineage tracing experiments. Black bars indicate tamoxifen dosing, and arrows indicate dates of removal. c-e, Right lateral views of early hearts show zones with heavy pre-artery specification. f, g, i, Low magnification of right lateral views

(leftmost panels) at late embryonic stages show labelling in the main right coronary artery; right panels focus on more distal branches of the right coronary artery. Table summarizes labelling and results. $n \geq 3$ hearts for each experiment. **h**, Quantification of *ApjCreER* and *Cx40CreER* lineage labelling indicates that most of the embryonic coronary artery is formed by cells specified within the E12.5–E14.5 time window. n = 3 for each *Cre*. Data shown as mean \pm s.d. Tam, tamoxifen. Scale bars, 100 μ m.

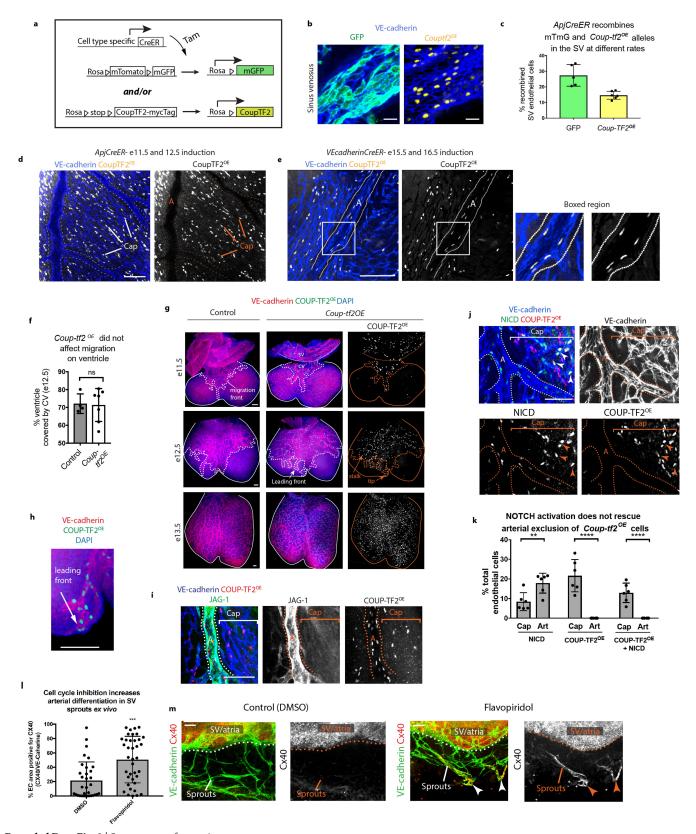




Extended Data Fig. 8 | Gene expression curves in E12.5 cells.

a, Expression of genes from the indicated categories along the SV–CV plexus–arterial differentiation continuum. The *x*-axis has individual cells organized as shown in Fig. 3a, and gene expression is plotted as LOESS curves. Raw data points are shown as dots. **b**, GSEA; cell cycle pathways

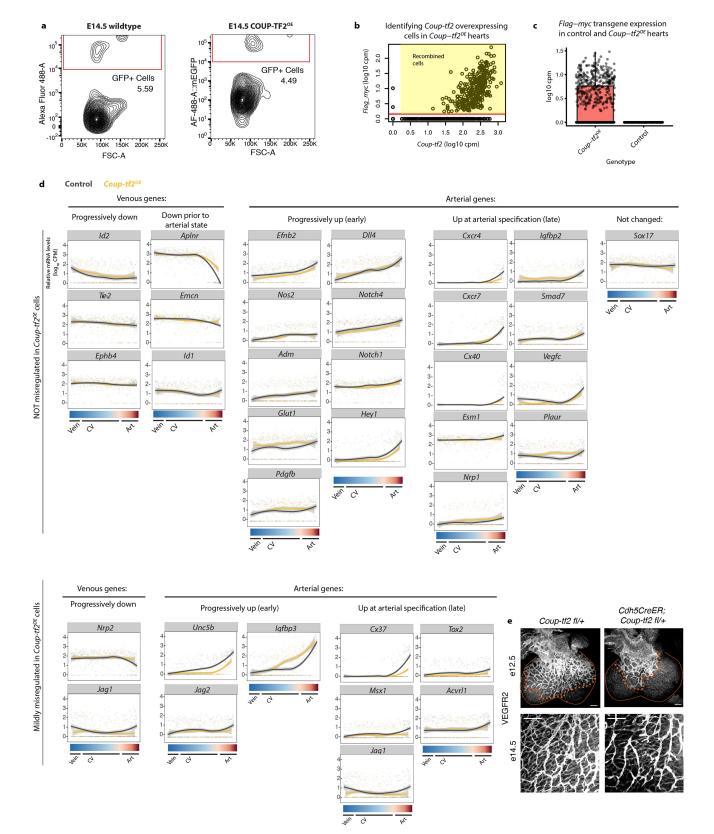
shown in bold. **c**, Pre-artery cells $(Cx40^+Cxcr4^+Apj^-)$ segregate to the non-cycling quadrant of rPCA plots. Arrows indicate cell cycle progression. **d**, Quantification of EdU labelling in pre-artery cells. Data shown as mean \pm s.d. n=6 hearts. *** $P \le 0.001$. Unpaired two-tailed t-test was used to calculate P value.



Extended Data Fig. 9 \mid See next page for caption.

Extended Data Fig. 9 | Effect of COUP-TF2 overexpression during coronary vessel development. a, Schematic of transgenes used to study Coup-tf2 overexpression in coronary cells. b, c, Recombination is not complete in the SV with tamoxifen at E9.5 and E10.5 as shown in whole-mount confocal images (b) and quantification (c). Control GFP is visualized by direct fluorescence, and COUP-TF2^{OE} through immunostaining for the myc tag. For c, ApjCreER Rosa^{mTmG}, n = 5 hearts. ApjCreER Coup-tf2^{OE}, n = 6 hearts. **d**, Tamoxifen dosing at E11.5 and E12.5 fills capillaries with recombined cells, but still resulted in $Coup-tf^{2OE}$ cells being excluded from arteries (A). e, Induction of $Coup-tf^{2OE}$ throughout vasculature shows that overexpressing cells can exist in arteries. f, Quantification of ventricle coverage at E12.5. n = 4 control hearts, n = 7 COUP-TF2^{OE} hearts. ns, P > 0.05. P = 0.8868. g, Wholemount confocal images of control and Coup-tf2^{OE} hearts at different stages of development. Coronary migration (dotted line) on the dorsal side of the ventricle (outlined with solid line) is similar in both genotypes. **h**, High magnification of E12.5 Coup-tf2^{OE} heart shown in **g** highlights the positioning of transgenic cells at both the leading front and trailing cells. i, COUP-TF2^{OE} cells can become part of the JAG-1-positive artery if induced after pre-artery specification with Cx40CreER. j, Mosaic

experiment in which constitutive expression of the NOTCH intracellular domain (NICD) is induced at the same time as *Coup-tf2*^{OE}. This manipulation creates a vasculature containing three different transgene combinations: 1. NICD; 2. COUP-TF2^{OE}; or 3. NICD + COUP-TF2^O (arrowheads). Those containing just the NICD (category 1) are the only transgenic cells that contribute to arterial vessels. k, Quantification of the percentage of endothelial cells in capillaries and arteries (Art) with the three transgenic combinations. NICD-expressing cells preferred arteries whereas COUP-TF2OE cells avoid arteries, the latter of which was not rescued by NICD. n = 6 hearts. ** $P \le 0.01$; **** $P \le 0.0001$. For NICD capillary versus artery, P = 0.0070. For COUP-TF2^{OE} capillary versus artery, $P = 7.49224 \times 10^{-5}$. For COUP TF2^{OE} + NICD capillary versus artery, $P = 8.07734 \times 10^{-5}$. I, The CDK inhibitor flavopiridol increased arterial specification (Cx40) in an SV sprouting assay. n = 33 control explants, n = 38 treated explants. *** $P \le 0.001$. m, Immunostaining of endothelial sprouts (VE-cadherin⁺) migrating from SV or atria tissue explants with Cx40 showed the increase in this arterial marker (arrowheads) with flavopiridol treatment. Data shown as mean \pm s.d. A two-tailed unpaired *t*-test was performed to determine *P* values. Scale bars, **b**, 20 μ m; **d**, **e**, **g**–**j**, 100 μ m; **m**, 25 μ m.



Extended Data Fig. 10 | Gene expression curves in E14.5 control and Coup- $tf2^{OE}$ cells. a, FACS plots of the GFP-marked cells from control and Coup- $tf2^{OE}$ hearts that were processed for scRNA-seq. b, Criteria for identifying Coup- $tf2^{OE}$ cells was >1 read of the flag and myc sequences included in the transgene. c, Comparing the number of flag and myc reads in control and Coup- $tf2^{OE}$ hearts confirms the specificity of this parameter for transgenic cells. Control: n = 409 cells. COUP-TF2^{OE}: n = 714 cells. The centre line corresponds to the median; the upper and lower hinges correspond to the first and third quartile, respectively; the whiskers extend

to the largest value or to 1.5 \times IQR (inter-quartile range, or distance between quartiles), whichever is smaller. **d**, Expression of genes from the indicated categories along the vein–CV plexus–arterial axis. The *x*-axis has individual cells organized as shown in Fig. 5b. Lines are LOESS curves of gene expression and raw data points are shown as dots. Shaded region represents the 95% confidence interval of the LOESS curve. **e**, Hypoplastic coronary vasculature with heterozygous deletion of *Coup-tf2* in endothelial cells. Scale bars, 100 μ m.

ARTICLE

IL-23 secreted by myeloid cells drives castration-resistant prostate cancer

Arianna Calcinotto¹, Clarissa Spataro¹, Elena Zagato¹, Diletta Di Mitri¹, Veronica Gil², Mateus Crespo², Gaston De Bernardis¹, Marco Losa¹, Michela Mirenda¹, Emiliano Pasquini¹, Andrea Rinaldi¹, Semini Sumanasuriya², Maryou B. Lambros², Antje Neeb², Roberta Lucianò³, Carlo A. Bravi³, Daniel Nava-Rodrigues², David Dolling², Tommaso Prayer-Galetti⁴, Ana Ferreira², Alberto Briganti³, Antonio Esposito⁵, Simon Barry⁶, Wei Yuan², Adam Sharp², Johann de Bono² & Andrea Alimonti^{1,7,8,9}*

Patients with prostate cancer frequently show resistance to androgen-deprivation therapy, a condition known as castration-resistant prostate cancer (CRPC). Acquiring a better understanding of the mechanisms that control the development of CRPC remains an unmet clinical need. The well-established dependency of cancer cells on the tumour microenvironment indicates that the microenvironment might control the emergence of CRPC. Here we identify IL-23 produced by myeloid-derived suppressor cells (MDSCs) as a driver of CRPC in mice and patients with CRPC. Mechanistically, IL-23 secreted by MDSCs can activate the androgen receptor pathway in prostate tumour cells, promoting cell survival and proliferation in androgen-deprived conditions. Intra-tumour MDSC infiltration and IL-23 concentration are increased in blood and tumour samples from patients with CRPC. Antibody-mediated inactivation of IL-23 restored sensitivity to androgen-deprivation therapy in mice. Taken together, these results reveal that MDSCs promote CRPC by acting in a non-cell autonomous manner. Treatments that block IL-23 can oppose MDSC-mediated resistance to castration in prostate cancer and synergize with standard therapies.

Prostate cancer is the most commonly diagnosed cancer in males in the world and the second leading cause of mortality in males that is attributable to cancer¹. After it was shown that androgens and androgen receptor (AR) signalling promote prostate cancer progression, androgendeprivation therapy (ADT) has become the main prostate cancer therapy for patients at different stages of disease². However, a considerable fraction of patients receiving such treatments ultimately progress to a more aggressive disease, developing CRPC². The prognosis for patients with CRPC remains poor and the treatment of these patients remains a major unmet medical need²⁻⁸. A better understanding of the mechanisms that drive CRPC could identify more effective therapies. Deregulated AR signalling, induced by genomic amplification of the AR locus, AR splice variants and activation of co-regulators of the AR, is considered the major determinant of CRPC². Activation of several AR-alternative signalling pathways also promotes CRPC^{9,10}. However, these mechanisms suggest that cell-autonomous alterations occur in prostate tumour cells and do not take into consideration that these cells are surrounded by a complex tumour microenvironment. The wellestablished dependency of cancer cells on the tumour microenvironment¹¹ suggests that the non-cancer-cell component of the tumour may control prostate cancer progression, although the contribution of the tumour microenvironment, and in particular of the tumour immune response to the emergence of CRPC, remains unknown 12,13 . We and others have previously reported that MDSCs are a prominent immune cell subset infiltrating the CRPC microenvironment^{14–16}. MDSCs are a heterogeneous population of activated immune cells that are expanded in pathological conditions, including cancer, and that have potent immunosuppressing activity¹⁷. On the basis of their expression markers, MDSCs can be classified into monocytic MDSCs or polymorphonuclear (PMN)-MDSCs¹⁸. Higher numbers of circulating and tumour-infiltrating MDSCs have been observed in a large fraction of patients who have different types of tumours including prostate cancer^{17,19,20}. MDSCs can support tumorigenesis by either suppressing the antitumour immune response or by promoting angiogenesis and senescence evasion in a number of contexts including prostate cancer^{15,16,21}. MDSCs have been also found to be increased in patients that do not respond to ADT²⁰. However, whether MDSCs support androgen-independent tumour growth and the emergence of CRPC remains unknown. Here, we show that IL-23 secreted by increased numbers of MDSCs in both human and mouse prostate tumours can confer androgen independence in a non-cell autonomous manner through the activation of AR signalling. Inhibition of IL-23 or IL-23 receptor signalling in these tumours restores sensitivity to ADT.

MDSCs confer castration resistance

By analysing biopsies from patients with castration-sensitive prostate cancer (CSPC) and CRPC, we found that PMN-MDSCs (CD11b⁺CD33⁺ CD15⁺ cells)¹⁸ were enriched in CRPC and localized in close proximity to EpCAM⁺ epithelial tumour cells (Fig. 1a and Extended Data Fig. 1a). Notably, increased PMN-MDSCs in tumours were not associated with increased levels of CD11b⁺CD15⁻ cells (Extended Data Fig. 1b). Therefore, we hypothesized that tumour infiltrating PMN-MDSCs could directly contribute to the emergence of CRPC. We investigated this hypothesis using the *Pten*-null prostate conditional knockout ($Pten^{PC-/-}$) mouse model and two additional allograft models of prostate cancer (TRAMP-C1 and MyC-CaP) that develop tumours driven by *Pten* loss, p53 and RB inactivation, and MYC amplification, respectively². As previously reported²², surgical castration of $Pten^{PC-/-}$ mice leads to initial tumour regression (castration-sensitive phase, t=4 weeks) followed by tumour progression and emergence

¹Institute of Oncology Research (IOR), Oncology Institute of Southern Switzerland, Bellinzona, Switzerland. ²The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, UK. ³Division of Oncology, Unit of Urology, URI, IRCCS Ospedale San Raffaele, Milan, Italy. ⁴Department of Urology, University of Padova, Padova, Italy. ⁵Experimental Imaging Center, San Raffaele Scientific Institute, Milan, Italy. ⁵IMED Oncology AstraZeneca, Li Ka Shing Centre, Cambridge, UK. ¹Università della Svizzera italiana, Faculty of Biology and Medicine, University of Lausanne, Switzerland. ⁵Department of Medicine, Venetian Institute of Molecular Medicine, University of Padova, Padova, Italy. *e-mail: andrea.alimonti@ior.iosi.ch

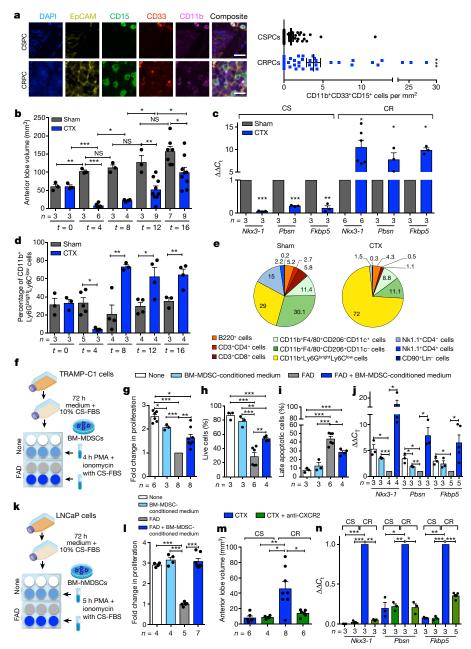


Fig. 1 | MDSCs infiltrate CRPC paralleling the activation of AR pathway and conferring resistance to castration to prostate cancer. a, CD11b+CD33+CD15+ PMN-MDSCs within the tumours of patients with CSPC or CRPC. Left, representative images of CSPC and CRPC from patient 1. Right, Quantification. EpCAM, yellow; CD15, green; CD33, red; CD11b, pink; DAPI, blue. n = 51 biologically independent patients per group reported; data are mean \pm s.e.m. Statistical analyses (negative binomial regression model): P < 0.001. **b-d**, $Pten^{PC-/-}$ mice that were sham-operated (Sham) or surgically castrated (CTX) $Pten^{PC-/-}$ mice at different time points. **b**, Tumour volume of the anterior prostate lobe. c, Quantitative PCR with reverse transcription (qRT-PCR) analyses of the indicated genes in the prostate tumours at t = 4 weeks (castrationsensitive phase (CS)) and t = 12 weeks (castration-resistant phase (CR)). **d**, Flow cytometry for tumour PMN-MDSCs (gated on CD45⁺ cells). e, Percentages of tumour-infiltrating immune cell populations (gated on CD45⁺ cells). f, Experimental schematic. CS-FBS, charcoal-stripped

of castration-resistant prostate tumours (castration-resistant phase, $t\!=\!12$ weeks) (Fig. 1b and Extended Data Fig. 1c, d). AR target genes are downregulated in tumours in the castration-sensitive phase and upregulated in the castration-resistant phase in castrated mice compared to sham-operated mice (Fig. 1c). To assess whether castration

affects the recruitment of PMN-MDSCs in these tumours, we measured the frequency of PMN-MDSCs (CD11b $^+$ Ly6G bright Ly6C low cells) 18 in sham-operated and castrated $Pten^{PC-/-}$ mice in a time course experiment. Notably, PMN-MDSCs number increased over time, paralleling the emergence of CRPC (Fig. 1b, d and Extended Data Figs. 1e, 2a).

Of note, PMN-MDSCs represented the major subset of immune cells that increased in *Pten*-null tumours upon castration (Fig. 1e and Extended Data Fig. 2b, c). This increase in PMN-MDSCs was validated in TRAMP-C1 and MyC-CaP castrated mice that develop CRPC within 10 days after castration (Extended Data Figs. 1f–i, 2a). Whereas PMN-MDSCs increased in castrated tumours, the frequency of tumour-infiltrating macrophages (TAMs) decreased (Fig. 1e and Extended Data Fig. 2c).

To assess whether factors secreted by MDSCs affect AR pathway signalling in prostate cancer cells, we co-cultured two mouse androgendependent prostate cancer cell lines, TRAMP-C1 and MyC-CaP, in the presence of conditioned medium obtained from bone marrow (BM)-derived MDSCs (Fig. 1f and Extended Data Fig. 3a, b). Notably, the conditioned medium of MDSCs sustained the proliferation and survival of cells cultured under full androgen deprivation (FAD), enhancing the transcription of AR target genes (Fig. 1g-j and Extended Data Fig. 3c-f). These results were further validated in both androgendependent (LNCaP and VCaP) and androgen-independent (22Rv1 and PC3) human prostate cancer cell lines cultured in the presence of conditioned medium from human BM-MDSCs (Fig. 1k, 1 and Extended Data Fig. 3g-i). Taken together, these data demonstrated that MDSCs can regulate, in a paracrine manner, androgen-deprivation sensitivity in prostate tumour cells. We next assessed whether depletion of MDSCs could delay the emergence of CRPC in castrated mice. We therefore treated castrated Pten^{PC-/-} mice, TRAMP-C1 and MyC-CaP allograft mice with AZD5069, a selective CXCR2 antagonist that is under clinical evaluation (Clinical Trial NCT03177187, https://clinicaltrials.gov/ct2/ show/NCT03177187). Treatment with the CXCR2 antagonist strongly reduced the tumour infiltration of PMN-MDSCs in all of the mouse models that were analysed (Extended Data Figs. 3j 4a, e). Notably, whereas Pten^{PC-/-} castrated mice treated with the CXCR2 antagonist did not progress to CRPC, untreated mice developed CRPC four months after castration as demonstrated by the levels of AR target genes (Fig. 1m, n). This finding was also confirmed in TRAMP-C1 and MyC-CaP allograft mice, in which inhibition of MDSC recruitment in the tumour delayed the emergence of CRPC as shown by decreased tumour size and level of AR target genes in treated mice, resulting in longer survival in mice that were treated with the CXCR2 antagonist (Extended Data Fig. 4a-h). Of note, treatment with the CXCR2 antagonist did not directly affect cell proliferation and AR activity in mouse prostate tumour cells cultured in vitro in FAD (Extended Data Fig. 3k, l). To corroborate the role of MDSCs as drivers of CRPC in the human setting, we co-injected human BM-MDSCs with LNCaP cells in NOD/ SCID mice, and assessed tumour growth over time in intact versus castrated hosts. In line with our previous results, the co-injection of LNCaP with human BM-MDSCs conferred resistance to castration (Extended Data Fig. 4i, j). Taken together, these data indicated that MDSCs are increased in CRPC and can promote proliferation of prostate tumour cells by sustaining AR signalling following androgen deprivation.

IL-23 drives insensitivity to androgen deprivation

To determine which MDSC-secreted factors drive castration resistance, we performed a NanoString nCounter gene expression assay in $Pten^{PC-/-}$ tumours from sham and castrated mice. IL-23 and one of the subunits of IL-23 receptor (IL12R β 1) were the most upregulated genes in tumours from castrate mice compared to controls (Extended Data Fig. 5a). Of note, factors that had previously been linked to the regulation of AR pathway, such as IL-6, were not upregulated in $Pten^{PC-/-}$ tumours after castration¹³ (Extended Data Fig. 5a). In line with this evidence, cytokine profile analysis of conditioned medium from mouse MDSCs showed that IL-23 was the most overexpressed factor produced by these cells (Extended Data Fig. 5b). Immunofluorescence and flow cytometry analyses further confirmed that tumour-infiltrating MDSCs expressed IL-23 in vivo, with PMN-MDSCs that infiltrated castration-resistant tumours expressing even higher levels of IL-23 compared to treatment-naive tumours (Fig. 2a, b). Moreover, similar to the results

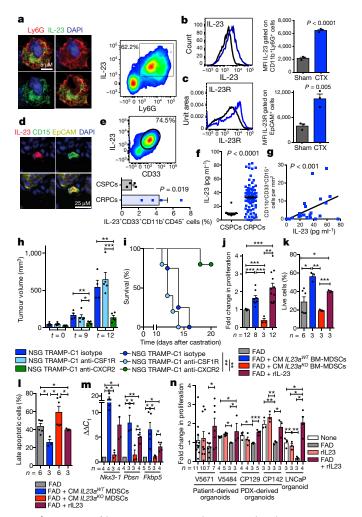


Fig. 2 | Tumour-infiltrating MDSCs produce IL-23 that drives insensitivity to androgen deprivation. a-c, Prostate tumours from $Pten^{PC-/-}$ sham-operated or CTX mice analysed at t = 12 weeks. a, Representative images of Ly6G⁺IL-23⁺ cells (nuclei, blue) and representative dot plot of Ly6G⁺IL-23⁺ cells gated on CD45⁺ cells in CTX mice. **b**, **c**, Representative histograms (left) and quantification (right; mean \pm s.e.m.) showing the mean fluorescence intensity (MFI) of IL-23 in CD45⁺CD11b⁺Ly6G⁺ cells (b) and MFI of IL-23R gated on CD45⁻EpCAM⁺ cells (c). n = 3 biologically independent mice per group. d, Representative images of IL-23⁺, CD15⁺, EpCAM⁺ cells within the tumours of patients with CRPC. a, d, Data were validated in at least three experiments. e, CD33⁺IL-23⁺CD11b⁺CD45⁺ cells within the tumours of patients with CSPC or CRPC (n = 4 biologically independent patients per cohort). f, IL-23 levels in the plasma of patients with CSPC (n = 20) and CRPC (n = 120). **g**, Correlation analyses of the numbers of tumourinfiltrating PMN-MDSCs and plasmatic IL-23 levels in patients with CRPC (n = 28). Statistical analyses (negative binomial regression model): P < 0.001. **h**, **i**, Tumour progression (t = days after castration) of NSG mice carrying TRAMP-C1 allografts treated with isotype control (untreated; n = 4), anti-CSF1R antibody (n = 5) or CXCR2 antagonist (n = 5). h, Average tumour volume. i, Survival curves reported as a Kaplan-Meier plot. Statistical analyses (log-rank (Mantel-Cox) test): **P < 0.01. j, TRAMP-C1 cell proliferation. k, Percentage of annexin V⁻ and 7AAD⁻ TRAMP-C1 cells. I, Percentage of annexin V⁺ and 7AAD⁺ TRAMP-C1 cells. m, qRT-PCR analyses of the indicated genes in TRAMP-C1 cells. n, Cell proliferation of 3D cultures of reported organoids. Recombinant (r) IL-23 conditions were normalized to the none or FAD treatment. e, f, h, j-n, Data are mean \pm s.e.m. Specific *n* values of biologically independent samples are shown in j-n. b, c, e, f, h, j-l, n, Statistical analyses (unpaired two-sided Student's t-test). m, Statistical analyses (paired two-sided Student's *t*-test): *P < 0.05; **P < 0.01; ***P < 0.001.

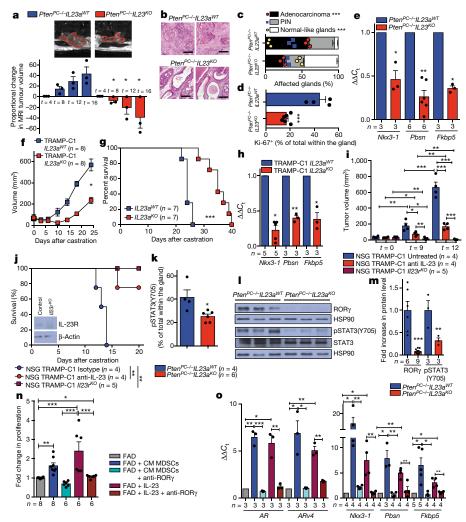


Fig. 3 | IL-23-IL-23R signalling axis regulates resistance to castration in prostate cancer in vivo and in vitro. a, Magnetic resonance imaging scans of representative Pten^{PC-/-}Il23a^{WT} and Pten^{PC-/-}Il23a^{KO} mice at the endpoint (top). Waterfall plot depicting proportional change in tumour response for $Pten^{PC-/-}Il23a^{WT}$ (n=3) and $Pten^{PC-/-}Il23a^{KO}$ (n=3)mice (bottom). Data are mean \pm s.e.m. Statistical analyses: unpaired two-sided Student's *t*-test: *P < 0.05; one-way ANOVA: P = 0.0008. **b**, Representative haematoxylin and eosin staining at the endpoint. Scale bars, 100 µm. Data are representative of at least three biologically independent mice. c, Quantification of adenocarcinoma, prostatic intraepithelial neoplasia (PIN) or normal-like glands in $\hat{Pten}^{PC-/-}Il23a^{WT}$ (n=3) and $Pten^{PC-/-}Il23a^{KO}$ (n=3) mice. **d**, Quantification of Ki-67 cells in $Pten^{PC-/-}Il23a^{WT}$ (n=4) and $Pten^{PC-/-}Il23a^{KO}$ (n=6) mice. e, qRT-PCR analyses of the prostate tumours. f, g, Tumour volume and survival curves of TRAMP-C1 *Il23a^{WT}* and TRAMP-C1 *Il23a^{KO}* mice. f, Statistical analyses (unpaired two-sided Student's t-test followed by Wilcoxon signed-rank test): *P < 0.05. **h**, qRT-PCR analyses in the tumours of TRAMP-C1 Il23aWT or TRAMP-C1 Il23aKO mice.

i, j, NSG males challenged with TRAMP-C1 or TRAMP-C1 $Il23r^{KO}$ cells after CTX were treated with isotype control (untreated) or anti-IL-23 antibody and monitored for tumour progression (i) and survival (j). Western blot of IL-23R in TRAMP-C1 (control) or TRAMP-C1 *Il23r*^{KO} cells is shown in the inset (performed at least twice). g, j, Statistical analyses (log-rank (Mantel-Cox) test): **P < 0.01; ***P < 0.001. **k**, Quantification of pSTAT3(Y705)⁺ cells in $Pten^{PC-/-}Il23a^{WT}$ (n=4) and $Pten^{PC-/-}Il23a^{KO}$ (n=6) mice. **l**, **m**, Western blot (**l**) and quantification (**m**) of RORγ, pSTAT3(Y705) and total STAT3 levels in prostate tumours. n, TRAMP-C1 cell proliferation. o, qRT-PCR analyses in TRAMP-C1 cells. c, d, k, Data are mean \pm s.e.m. of one tumour per mouse (mean of three sections per tumour, ≥ 3 fields per section). **e-k**, **m-o**, Data are reported as mean \pm s.e.m. Specific *n* values of biologically independent mice (i-k) and biological independent samples (m-o) are shown. c, d, i, k, m, n, Statistical analyses (unpaired two-sided Student's t-test). **e**, **h**, **o**, Statistical analyses (paired two-sided Student's t-test): *P < 0.05; **P < 0.01; ***P < 0.001.

found in mice, PMN-MDSCs that infiltrated tumour biopsies from patients with CRPC expressed IL-23 (Fig. 2d, e and Extended Data Fig. 6a). In addition, the frequency of IL-23-producing tumour-infiltrating PMN-MDSCs was higher in CRPC biopsies than in CSPC biopsies (Fig. 2e). Notably, expression of CXCL5, a chemokine that stimulates chemotaxis of myeloid cells through CXCR2²³, was strongly upregulated in castrated tumours compared to controls (Extended Data Fig. 5a, c, d). This, together with the finding that CXCR2 inhibition efficiently decreases the recruitment of MDSCs in castrated mice, indicates that CXCL5 is a major regulator of MDSC recruitment in CRPC.

We next assessed the levels of the IL-23 receptor (IL-23R) in tumours from sham and castrated $Pten^{PC-/-}$ mice, and found that IL-23R levels

increased in tumour cells following castration (Fig. 2c). This was further validated in TRAMP-C1 cells cultured in androgen-deprived conditions in vitro (Extended Data Fig. 5e, f). Furthermore, plasma levels of IL-23 in patients with CRPC were substantially higher than in patients with CSPC (Fig. 2f and Extended Data Fig. 5g) and statistically correlated with tumour-infiltrating PMN-MDSC counts (EpCAM⁻CD11b⁺CD33⁺CD15⁺ cells; Fig. 2g) but not with other myeloid cell population counts (CD11b⁺CD15⁻ cells; Extended Data Fig. 6b). Overall, these data demonstrate that IL-23 is increased in both mouse and human CRPCs, with IL-23 levels correlating with the number of tumour-infiltrating PMN-MDSCs. Overall, tumour biopsies from patients with CRPC also had much higher *IL23A* (which

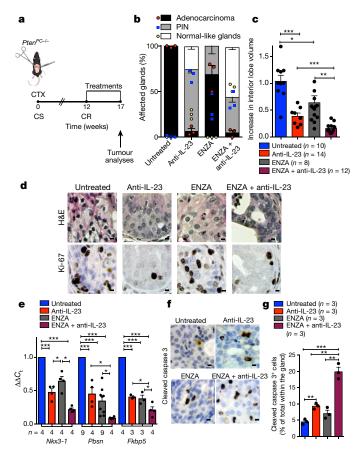


Fig. 4 | IL-23 inhibition improves ENZA efficacy in vivo.

a, *Pten*^{PC-/-} mice with castration-resistant prostate cancer (12 weeks after castration) were randomly assigned to treatments in a preclinical study. Treatments: isotype control (untreated), anti-IL-23 antibody (100 ng per mouse injected intraperitoneally weekly), ENZA (30 mg kg⁻¹ per day administered daily by oral gavage on a Monday to Friday schedule) and ENZA in combination with anti-IL-23 antibody (ENZA + anti-IL-23). **b**, Histological score. n = 3 biologically independent mice. Statistical analyses (two-way ANOVA): P < 0.001. c, Fold increase in the volume of the anterior lobe of the prostate (fold change over the untreated group). d, Representative haematoxylin and eosin (H&E) and Ki-67 staining in the tumours at completion of the study. Scale bars, $50 \,\mu\text{m}$. **e**, qRT–PCR analyses of the indicated genes in the prostate tumours of CTX Pten^{PC-/} mice at completion of the preclinical study. Statistical analyses (two-sided paired Student's *t*-test): *P< 0.05; ***P< 0.001. **f**, Representative cleavedcaspase3 staining in the tumours after one week of treatments. Scale bars, 50 μm. g, Quantification of cleaved-caspase 3 (percentage of the total number of cells within the glands). d, f, Data are representative of at least three biologically independent mice. **b**, **g**, Data are mean \pm s.e.m. of one tumour per mouse (mean of three sections per tumour, three or more fields per section). **b**, **c**, **e**, **g**, Data are mean \pm s.e.m. Specific *n* values of biological independent mice are shown in c, e, g. c, g, Statistical analyses (unpaired two-sided Student's t-test): NS, not significant; *P < 0.05; **P < 0.01; ***P < 0.001.

encodes IL-23) and *IL23R* mRNA levels than treatment-naive patients (Extended Data Fig. 6c, d) and *IL23A* mRNA levels were linked to MDSC-associated mRNA levels in biopsies from patients with CRPC (Extended Data Fig. 6e, f).

In line with this evidence, the primary source of IL-23 in tumours from castrated *Pten*^{PC-/-} mice was the population of PMN-MDSCs (Extended Data Fig. 6g). Note that TAMs and epithelial tumour cells were only a small fraction of the total IL-23⁺ cells in these tumours (Extended Data Fig. 6g). In keeping with this, IL-23 levels significantly decreased in tumours from castrated *Pten*^{PC-/-} mice that were depleted of MDSCs (Extended Data Fig. 6h), whereas the depletion of TAMs by an CSF1R antibody²⁴ in mice with TRAMP-C1 allografts did not delay

the emergence of CRPC. Conversely, the reduction of PMN-MDSCs by a CXCR2 antagonist robustly delayed the emergence of CRPC (Fig. 2h, i and Extended Data Fig. 6i). Overall, therefore, IL-23 levels significantly decreased in tumours depleted of MDSCs but not of TAMs (Extended Data Fig. 6j).

To functionally validate these findings, we cultured prostate tumour cells in the presence of conditioned medium from BM-MDSCs from Il23a wild-type (IL23WT BM-MDSCs) or Il23a knockout mice (IL23ko BM-MDSCs). The conditioned medium of *IL23*^{WT} BM-MDSCs, as well as treatment with recombinant IL-23, promoted the proliferation, survival and increased the transcription of AR target genes in prostate cancer cells kept in FAD, whereas the conditioned medium of IL23^{KO} BM-MDSCs was unable to affect these parameters (Fig. 2j-m). Of note, deleting IL-23 in BM-MDSCs did not affect the levels of other secreted factors in these cells (Extended Data Fig. 7a). Indeed, IL23WT and IL23KO BM-MDSCs had equal immunosuppressive capabilities (Extended Data Fig. 7b). These results were further validated in a subset of androgen-dependent organoids that were derived from patient-derived xenografts and LNCaP cells kept in FAD and treated in the presence or absence of human recombinant IL-23 (Fig. 2n). Taken together, these findings identify IL-23 as an MDSC-secreted factor that can sustain the proliferation and survival of prostate cancer cells as well as the transcription of AR-driven target genes in prostate cancer cells cultured in FAD.

MDSCs activate the IL-23-RORγ pathway

To determine whether MDSC-derived IL-23 promotes the emergence of resistance to castration in prostate cancers in vivo, we next reconstituted lethally irradiated sham-operated or castrated *Pten*^{PC-/-} mice with bone marrow precursors from $Il23a^{WT}$ or $Il23a^{KO}$ mice (yielding Pten^{PC-/-}IL23^{WT} and Pten^{PC-/-}Il23a^{KO} mice, respectively; Extended Data Fig. 8a). Mice were reconstituted with bone marrow that was deprived of T, B and natural killer (NK) cells (Extended Data Fig. 8b). The absence of IL-23 in the myeloid compartment led to a marked reduction in the tumour volume of the prostate cancers specifically in castrated PtenPC-/- mice (Fig. 3a and Extended Data Fig. 8c), with normalization of glands that are affected by prostate cancer and a major reduction in Ki-67⁺ cells (Fig. 3b-d and Extended Data Fig. 8d). Notably, AR target genes were robustly downregulated in prostate tumours from $Pten^{PC-/-}Il23a^{KO}$ compared to tumours from Pten^{PC-/-}Il23a^{WT} mice (Fig. 3e). These data were then also validated in TRAMP-C1-allograft irradiated mice reconstituted with Il23aWT and Il23a^{KO} bone marrow precursors (yielding TRAMP-C1 Il23a^{WT} and TRAMP-C1 Il23aKO mice; Extended Data Fig. 9a). In TRAMP-C1 *Il23a*^{KO} mice, the absence of IL-23 in the myeloid compartment significantly delayed the emergence of CRPC as demonstrated by decreased tumour size and tumour cell proliferation as well as reduced AR-driven target gene expression and significantly increased survival of the TRAMP-C1 *Il23a*^{KO} mice (Fig. 3f-h and Extended Data Fig. 9b, c). Reconstitution with *Il23a^{KO}* bone marrow did not alter the recruitment of MDSCs into the tumours and spleens of reconstituted mice (Extended Data Fig. 9d, e). Of note, treatment with anti-IL-23 antibodies or genetic inactivation of IL-23R in NOD/SCID/ γ (NSG) TRAMP-C1 allografts confirmed these results (Fig. 3i, j), demonstrating that IL-23 directly promotes resistance to castration in prostate cancer by binding to IL-23R that are present on tumour cells.

IL-23 has been reported to regulate the activation of STAT3–ROR γ expression in naive CD4 T cells^{25–27}; both STAT3 and ROR γ can affect AR signalling in prostate cancer^{28,29}. We therefore evaluated whether IL-23 secreted by MDSCs affected the STAT3–ROR γ signalling axis in prostate cancer in a non-cell autonomous manner. Inactivation of IL-23 in the myeloid compartment of castrated $Pten^{PC-/-}$ mice significantly decreased overall tumour levels of phosphorylated (p)STAT3 and ROR γ in vivo (Fig. 3k–m and Extended Data Fig. 8d); this was also the case in the TRAMP-C1 model (Extended Data Fig. 9f–h). ROR γ inhibition in vitro also abrogated the proliferative advantage conferred by conditioned medium from MDSCs or IL-23 treatment in TRAMP-C1

cells kept in FAD, and inhibited the transcription of the full-length form of the AR and its constitutively active splice variant (ARv4) as well as downstream AR target genes (Fig. 3n, o). Taken together, these data demonstrate that IL-23 released by MDSCs into the tumour microenvironment acts directly on the pSTAT3–ROR γ signalling axis to drive the transcription of AR and its splice variant and downstream target genes, thus favouring the proliferation and survival of the prostate cancer cells in androgen-ablation conditions.

IL-23 targeting improves the efficacy of ADT

To evaluate the therapeutic relevance of our findings, we next assessed whether IL-23 inhibition by antibody blockade could reverse resistance to castration in prostate cancer in *Pten*^{PC-/-} mice. Because anti-IL-23 antibodies are currently being evaluated in clinical trials for the treatment of autoimmune diseases³⁰, including psoriasis and are clinically well-tolerated³¹, we treated mice carrying *Pten*^{PC-/-} tumours that had become resistant to surgical castration with an anti-IL-23 antibody in combination with the AR antagonist enzalutamide (ENZA; Fig. 4a). ENZA is a standard treatment for patients with CRPC after primary ADT^{2,32}. Our preclinical study showed that anti-IL-23 increased the efficacy of ENZA (Fig. 4b, c); in mice treated with anti-IL-23 and ENZA, we observed a normalization of prostate glands that were affected by cancer (Fig. 4b and Extended Data Fig. 10a), with decreased tumour volume (Fig. 4c) and proliferation (Extended Data Fig. 10a, b), whereas in mice that were treated with ENZA, alone the treatment was ineffective. Combined anti-IL-23 and ENZA were associated with a robust inhibition of the AR activity and induction of apoptosis of the tumour cells (Fig. 4e-g). Taken together, these data demonstrate that anti-IL-23 treatment can reverse resistance to castration in prostate cancer and enhance the efficacy of ENZA.

Discussion

Our study has identified IL-23 production by MDSCs as a driver of CRPC and adds novel mechanistic insights on how prostate cancers can become insensitive to androgen deprivation and AR blockade. We also report on a different role for MDSCs in cancer, describing an unexpected function for this immune subset. Previous data demonstrated that MDSCs can support tumorigenesis in many cancers through different mechanisms 15,17,18, with preclinical studies indicating that the inactivation of MDSCs increased immune-checkpoint blockade efficacy in CRPC models¹⁶. IL-23 has also been previously implicated in cancer progression in the context of a different tumour type as a regulator of the pro-tumour immune response^{33–35}. However, to our knowledge, the discovery described here, that IL-23 produced by MDSCs regulates resistance to castration in prostate cancer by sustaining AR signalling, was previously unknown, and adds novel mechanistic insights on how these immune cells support tumorigenesis. This work also shows that inhibition of IL-23 can reverse ADT resistance in men suffering from advanced prostate cancer (Extended Data Fig. 10c).

In conclusion, we describe an alternative immunotherapeutic strategy for treating advanced prostate cancer that, unlike most other treatments, is not focused on re-activating the function of cytotoxic T lymphocytes against tumour cells. Immunotherapeutic strategies that reactivate cytotoxic T cells by immune-checkpoint blockade have been, to date, only active against a small subset of prostate cancers that are characterized by DNA-repair defects and higher neoantigen loads with increased infiltration of T lymphocytes^{36,37}. Our results demonstrate, on the other hand, that MDSCs are a major player in the endocrine resistance in prostate cancer and that immunotherapies that target the blockade of either MDSC recruitment into tumour, or the direct inhibition of IL-23 can be effective therapeutic strategies for patients that have these lethal and common diseases. Because anti-IL-23 antibodies have been well-tolerated in clinical trials involving patients with autoimmune diseases³⁰, these deserve to be clinically evaluated in men that have lethal prostate cancer. We envision that this immunotherapeutic strategy targeting paracrine IL-23 in combination with established

endocrine anticancer treatments is highly likely to improve treatment outcome for this common male cancers.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0266-0.

Received: 25 July 2017; Accepted: 29 May 2018; Published online 27 June 2018.

- Torre, L. A. et al. Global cancer statistics, 2012. CA Cancer J. Clin. 65, 87–108 (2015).
- Watson, P. A., Arora, V. K. & Sawyers, C. L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* 15, 701–711 (2015).
- Bianchini, D. et al. Antitumour activity of enzalutamide (MDV3100) in patients with metastatic castration-resistant prostate cancer (CRPC) pre-treated with docetaxel and abiraterone. Eur. J. Cancer 50, 78–84 (2014).
- Zhang, T. et al. Exploring the clinical benefit of docetaxel or enzalutamide after disease progression during abiraterone acetate and prednisone treatment in men with metastatic castration-resistant prostate cancer. Clin. Genitourin. Cancer 13, 392–399 (2015).
- Badrising, S. et al. Clinical activity and tolerability of enzalutamide (MDV3100) in patients with metastatic, castration-resistant prostate cancer who progress after docetaxel and abiraterone treatment. Cancer 120, 968–975 (2014).
- Noonan, K. L. et al. Clinical activity of abiraterone acetate in patients with metastatic castration-resistant prostate cancer progressing after enzalutamide. *Ann. Oncol.* 24, 1802–1807 (2013).
- Schrader, A. J. et al. Enzalutamide in castration-resistant prostate cancer patients progressing after docetaxel and abiraterone. Eur. Urol. 65, 30–36 (2014).
- Sartor, O. & de Bono, J. S. Metastatic prostate cancer. N. Engl. J. Med. 378, 645–657 (2018).
- Karantanos, T., Ćorn, P. G. & Thompson, T. C. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene* 32, 5501–5511 (2013).
- Carver, B. S. et al. Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. Cancer Cell 19, 575–586 (2011).
- Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19, 1423–1437 (2013).
- Ammirante, M., Luo, J. L., Grivennikov, S., Nedospasov, S. & Karin, M. B-cellderived lymphotoxin promotes castration-resistant prostate cancer. *Nature* 464, 302–305 (2010).
- Lee, G. T. et al. Bone morphogenetic protein-6 induces castration resistance in prostate cancer cells through tumor infiltrating macrophages. Cancer Sci. 104, 1027–1032 (2013).
- Toso, A. et al. Enhancing chemotherapy efficacy in *Pten*-deficient prostate tumors by activating the senescence-associated antitumor immunity. *Cell Rep.* 9, 75–89 (2014).
- Di Mitri, D. et al. Tumour-infiltrating Gr-1⁺ myeloid cells antagonize senescence in cancer. Nature 515, 134–137 (2014).
- Lu, X. et al. Effective combinatorial immunotherapy for castration-resistant prostate cancer. Nature 543, 728–732 (2017).
- Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. Nat. Rev. Immunol. 9, 162–174 (2009).
- Bronte, V. et al. Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards. Nat. Commun. 7, 12150 (2016).
- Hossain, D. M. et al. TLR9-targeted STAT3 silencing abrogates immunosuppressive activity of myeloid-derived suppressor cells from prostate cancer patients. *Clin. Cancer Res.* 21, 3771–3782 (2015).
- Mehra, N. et al. Myeloid-derived suppressor cells (MDSCs) in metastatic castration-resistant prostate cancer (CRPC) patients (PTS). Ann. Oncol. 27, 757P (2016).
- Murdoch, C., Muthana, M., Coffelt, S. B. & Lewis, C. E. The role of myeloid cells in the promotion of tumour angiogenesis. *Nat. Rev. Cancer* 8, 618–631 (2008).
- Lunardi, A. et al. A co-clinical approach identifies mechanisms and potential therapies for androgen deprivation resistance in prostate cancer. *Nat. Genet.* 45, 747–755 (2013).
- Bezzi, M. et al. Diverse genetic-driven immune landscapes dictate tumor progression through distinct mechanisms. Nat. Med. 24, 165–175 (2018).
- 24. Gordon, S. R. et al. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature* **545**, 495–499 (2017).
- Durant, L. et al. Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity* 32, 605–615 (2010).
- Kastelein, R. A., Hunter, C. A. & Cua, D. J. Discovery and biology of IL-23 and IL-27: related but functionally distinct regulators of inflammation. *Annu. Rev. Immunol.* 25, 221–242 (2007).
- Zhou, L. et al. IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nat. Immunol.* 8, 967–974 (2007).
- 28. Chen, T., Wang, L. H. & Farrar, W. L. Interleukin 6 activates androgen receptormediated gene expression through a signal transducer and activator of



- transcription 3-dependent pathway in LNCaP prostate cancer cells. *Cancer Res.* **60**, 2132–2135 (2000).
- Campa, M., Mansouri, B., Warren, R. B. & Menter, A. A review of biologic therapies targeting IL-23 and IL-17 for use in moderate-to-severe plaque psoriasis. *Dermatol. Ther. (Heidelb.)* 6, 1–12 (2016).
- Gordon, K. B. et al. A phase 2 trial of guselkumab versus adalimumab for plaque psoriasis. N. Engl. J. Med. 373, 136–144 (2015).
- Moreira, R. B. et al. Differential side effects profile in patients with mCRPC treated with abiraterone or enzalutamide: a meta-analysis of randomized controlled trials. *Oncotarget* 8, 84572–84578 (2017).
- Gangemi, S. et al. Clinical significance of circulating interleukin-23 as a prognostic factor in breast cancer patients. J. Cell. Biochem. 113, 2122–2125 (2012).
- Grivennikov, S. I. et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. Nature 491, 254–258 (2012).
- Teng, M. W. L. et al. IL-23 suppresses innate immune response independently of IL-17A during carcinogenesis and metastasis. Proc. Natl Acad. Sci. USA 107, 8328–8333 (2010).
- 36. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
- 37. Graff, J. N. et al. Early evidence of anti-PD-1 activity in enzalutamide-resistant prostate cancer. *Oncotarget* **7**, 52810–52817 (2016).

Acknowledgements We acknowledge all members of the A.A. laboratory and M. Bellone for scientific discussions. R. D'Antuono helped with immunofluorescence acquisition. We thank F. Sallusto and R. Marzi for providing \$\mathref{I}23a^{\mathref{K}}\$^{\mathref{K}}\$ mice and all the members of the IRB animal core facility and E. Mira Catò for technical assistance and the animal work; T. Canu and A. Napolitano (Preclinical Imaging Facility of the San Raffaele Scientific Institute) for helping with MRI acquisition and analysis; and M. Montopoli (University) of Padova) for helping with sample logistics. This work was supported by ERC consolidator (683136) and Swiss Cancer League (KFS4267-08-2017) grant, Dr. Josef Steiner Foundation, Swiss Card-Onco-Grant of Alfred and Annemarie von Sick grant, and Helmut Horten Foundation, SNSF (310030_176045), PCUK (RIA15-ST2-018) and IBSA Foundation. Work in the de Bono laboratory was supported by the Movember Foundation, Prostate Cancer UK, the

US Department of Defense, the Prostate Cancer Foundation, Stand Up To Cancer, Cancer Research UK, and the UK Department of Health through an Experimental Cancer Medicine Centre grant. A.S. is supported by the Medical Research Council, the Academy of Medical Sciences and Prostate Cancer UK.

Reviewer information *Nature* thanks N. Brennen, D. Gabrilovich, M. Galsky and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.A. and A.C. conceived and designed the project. A.C., C.S., E.Z. and M.M. performed experiments. V.G. performed experiments with human organoids. D.D.M., M.C. and A.N. established and carried out fluorescence microscopy. A.R. checked the quality of the RNA and performed NanoString. D.D. performed statistical analyses of multiplex immunofluorescence in human tissue sections. W.Y. performed bioinformatic analyses. G.D.B. performed surgical castration or sham operation of all animals. M.L. performed immunohistochemical experiments and analysis. A.S., M.B.L., S.S., A.F., A.B., R.L., C.A.B. and T.P.-G. selected and provided human samples. D.N.-R. reviewed tissue segmentation, cell segmentation and phenotype determination maps of tissue sections. E.P. took care of genotyping the animals. S.B. provided the AZD5069 compound. A.E. performed MRI analyses and analysed the results. J.d.B. supervised human experiments and interpreted the data. A.C. and A.A. interpreted the data and wrote the paper.

Competing interests S.B. is affiliated to IMED Oncology AstraZeneca, Li Ka Shing Centre, Cambridge, UK and provided the AZD5069 compound. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0266-0.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0266-0.

Reprints and permissions information is available at http://www.nature.com/reprints

Correspondence and requests for materials should be addressed to A.A. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Animals. All mice were maintained under specific pathogen-free conditions in the IRB facility and experiments were performed according to national guidelines and regulations. All animal experiments were approved by the local ethics committee (TI13/2015 and TI04/2017). Male C57BL/6, FVB, NSG, NOD/SCID mice 6-8 weeks of age were purchased from Jackson Laboratories (Envigo) and acclimatized for at least a week before experiments. Male C57BL/6 IL-23p19KO (Il23aKO) mice27 were provided by F. Sallusto (IRB, Bellinzona) and used at 8 weeks of age. Male PtenPC-/- mice were generated and genotyped as previously described¹⁷. Female *Pten*^{loxP/loxP} mice were crossed with male PB^{cre4} transgenic mice and genotyped for Cre using the following primers: primer 1 (5'-AAAAGTTCCCCTGCTGATGATTTGT-3') and primer 2 (5'-TGTTTTTGACCAATTAAAGTAGGCTGTG-3') for $PTEN^{loxP/loxP}$; primer 1 (5'-TGATGGACATGTTCAGGGATC-3') and primer 2 (5'-CAGCCACC AGCTTGCATGA-3') for *PB*^{cre}. Surgical castration was performed under anaesthesia with isoflurane. Male *Pten*^{PC-/-} mice were 9–10 weeks old at the time of castration. Mice were monitored postoperatively for recovery from anaesthesia and checked daily for four days postoperatively. Surgical skin clips were removed on postoperative day 5. Mice undergoing treatment were administered control vehicle or therapeutic doses of the appropriate agents. Any mouse that showed signs of distress or lost more than 15% of their initial weight during treatment was euthanized by CO₂ asphyxiation. At the completion of study, mice were euthanized by CO₂ asphyxiation and tissue was collected for histology, mRNA and protein analysis and single-cell suspensions for flow cytometry. For allograft experiments, 2.5×10^6 TRAMP-C1 cells, 2.5×10^6 TRAMP-C1 $\emph{Il23r}^{KO}$ cells or 2×10^6 MyC-CaP cells were injected subcutaneously into the flank of male C57BL/6, C57BL/6 or FVB mice, respectively. For xenograft experiments 3×10^6 LNCaP cells were suspended with or without 3×10^6 human BM-MDSCs in a total volume of $100\,\mu l$ PBS and Matrigel (1:1) and implanted subcutaneously into the flank of NOD/SCID mice. When tumours were approximately 100 mm³, mice were randomized to the treatment groups. Tumour growth was monitored daily by measuring the tumour size with calliper. The tumour volume was estimated by calculating $4/3\pi(R_1 \times R_2 \times R_3)$, where R_1 and R_2 are the longitudinal and lateral radii and R_3 is the thickness of the tumour that protrudes from the surface of normal skin. Animals were euthanized when the tumour reached approximately 600 mm³. The local ethics committee approved the conduction of the in vivo experiments with maximal tumour sizes of $1,000 \, \text{mm}^3$.

Treatments. The CXCR2 antagonist (AZD5069; Astrazeneca) was administered with daily intraperitoneal injections at a final concentration of 100 mg kg $^{-1}$ on a Monday through Friday schedule. Control animals received vehicle. Enzalutamide (APExBio) was administered daily by oral gavage with a dose of 30 mg kg $^{-1}$ per day on a Monday through Friday schedule. Rat anti-IL-23 antibody (100 ng per mouse; G23-8; IgG1, kappa; eBioscience) or rat IgG1 isotype control (eBioscience) was administered weekly via intraperitoneal injection. For in vivo depletion of macrophages, mice were treated with 400 μ g anti-CSF1R (clone BE0213, BioXCell; on Mondays, Wednesdays and Fridays).

Cell lines. The TRAMP-C1, MYC-CaP, LNCaP, VCaP, 22Rv1 and PC3 cell lines were obtained from the ATCC and no other authentication method was performed. The TRAMP-C1 *Il23r^{KO}* cell line was generated in the laboratory with CRISPR-Cas9 methodology and authenticated by western blotting and FACS for the deletion of IL-23R. All cell lines were regularly tested for mycoplasma (MycoAlert Mycoplasma Detection kit).

Bone marrow reconstitution. Bone marrow was flushed from the femurs of male C57BL/6 or IL-23p19ko mice under sterile conditions with RPMI 1640 using a 21-gauge needle. Mononuclear cells were filtered, collected and checked for viability using trypan blue. Before transplantation, bone marrow-derived cells were depleted of CD3+ T cells, NK1.1+ NK cells and CD19+ B cells by magnetic bead separation (STEMCELL Technologies). Recipient C57BL/6 or $Pten^{PC-/-}$ mice were lethally irradiated (900 rad) and transplanted intravenously 2h later with 10^7 viable bone marrow cells from either C57BL/6 or IL-23p19ko mice. For TRAMP-C1 allografts, the animals were challenged subcutaneously with TRAMP-C1 cells upon bone marrow engraftment. When tumours reached approximately $100 \, \mathrm{mm}^3$, mice were surgically castrated and monitored for tumour progression.

Magnetic resonance imaging. Magnetic resonance imaging (MRI) was performed on castrated $Pten^{PC-/-}$ mice 0, 4, 8, 12 and 16 weeks after surgical castration or on CTX $Pten^{PC-/-}$ ll23 a^{WT} and CTX $Pten^{PC-/-}$ ll23 a^{WO} mice 4, 8, 12 and 16 weeks after surgical castration using a 7T preclinical magnetic resonance scanner (Bruker, BioSpec 70/30 USR, Paravision 5.1) equipped with 450/675 mT/m gradients (slew-rate: 3400–4500 T/m/s; rise-time 140 µs) and a mouse body volume coil. Mice were under general anaesthesia by 1.5–2% isoflurane vaporized in 100% oxygen (flow: 11 min⁻¹). Breathing and body temperature were monitored (SA Instruments, Inc.) and maintained around 30 breaths per minute and 37°C, respectively. MRI studies included a Rapid Acquisition with Relaxation Enhancement (RARE) High-Resolution T2-weighted (T2w) sequence with fat suppression

acquired in the axial plane (TR = 3,800 ms, TE = 45 ms, FOV = 27 mm \times 18 mm, spatial resolution = 0.094 \times 0.087 mm² per pixel, scan time = 8 min, thickness = 0.70 mm, 26 slices) and in the coronal plane (TR = 3,500 ms, TE = 38 ms, FOV = 33 mm \times 33 mm, spatial resolution = 0.129 \times 0,129 mm² per pixel, scan time = 5 min, thickness = 0.60 mm, 20 slices). Images were analysed using NIH software MIPAV (version 7.4.0). The circumference of the whole prostate was outlined on each RARE T2w axial slice containing identifiable prostate and the number of bounded pixels in each slice was computed and added to yield the prostate volume. Coronal T2w images were used for the accurate identification of the basal and apical limits of the prostate.

Differentiation of BM-MDSCs in vitro. Mouse BM-MDSCs were differentiated in vitro as previously described³⁸. In brief, bone marrow precursors were flushed from the femurs of C57BL/6 or IL-23p19ko mice with RPMI 1640 medium. The cell pellet was resuspended (one femur in 10 ml) in RPMI 1640 containing 10% heat-inactivated FBS and the cells were cultured in vitro in the presence of 40 ng ml⁻¹ GM-CSF and 40 ng ml⁻¹ IL-6. On day 4, the cells were washed and resuspended with RPMI 1640 containing 10% heat-inactivated charcoal-stripped FBS. The day after the cells were stimulated with PMA and ionomycin and after 4h the supernatant was collected and stored at $-80\,^{\circ}$ C. Analysis of soluble molecules was conducted with Mouse CytokineMAP B version 1.0 (Rules Based Medicine).

Human BM-MDSCs were differentiated in vitro by seeding 10^6 per ml bone marrow precursors in T25 flasks with RPMI 1640 containing 10% heat-inactivated FBS in the presence of 10 ng ml $^{-1}$ GM-CSF and 10 ng ml $^{-1}$ IL-6 for seven days³⁹. Complete medium was changed when required. After seven days, the cells were analysed by flow cytometry for CD11b, CD33, CD15, HLA-DR expression and when the CD11b+CD33+CD15+HLA-DR $^-$ population was higher than 80%, the cells were re-suspended in RPMI 1640 containing 10% heat-inactivated charcoal-stripped FBS and after one day stimulated with PMA and ionomycin for 5 h. The supernatant was then collected and stored at -80 °C.

In vitro culture experiments. Prostate cancer cell lines were starved in charcoal-stripped FBS medium for 72 h and then cultured with RPMI 1640 containing 10% heat-inactivated FBS (normal medium) or kept in full androgen-deprivation medium (FAD; RPMI 1640 containing 10% heat-inactivated charcoal-stripped FBS plus ENZA 10 μ M). Then, the cells were stimulated with or without conditioned medium obtained from activated BM-MDSCs, or recombinant IL-23 (100 ng ml $^{-1}$; R&D Systems), with or without ROR γ antibody (5 μ M; SR2211, Calbiochem). Then the cells were analysed using the crystal violet assay (after 72 h of culture, fold change over the FAD condition), stained with annexin V and 7AAD (after 72 h of culture) or collected for RNA extraction (after 24 h of culture; fold change over the FAD condition).

Analyses of IL23A and IL23R mRNA expression in clinical tumours. CSPC RNA-sequencing data for 550 patients was downloaded from the UCSC Cancer Browser (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/). Metastatic castration resistant prostate cancer (mCRPC) RNA-sequencing data for 122 mCRPC patients was generated as part of the SU2C effort⁴⁰. The paired-end transcriptome sequencing reads were aligned to the human reference genome (GRCh37/hg19) using TopHat2⁴¹ (version 2.1.0). Gene expression, as fragments per kilobase of exon per million fragments mapped (FPKM; normalized measure of gene expression), was calculated using Cufflinks⁴². MDSC marker (CD11b, CD33, CD14 and CD15) positive and negative was defined by the high and low quantiles of RNA expression of each transcript and IL23A and IL23R expression level in each biomarker groups were compared by Student's t-test. In order to compare gene expression level between TCGA and SU2C with minimum experimental bias, we included only genes expressed in both TCGA and SU2C with median expression level (FPKM) >0. The gene expression levels in each sample were quantile-normalized, and *IL23A* expression levels in CSPC and CRPC were compared using a Student's *t*-test. Human organoids. Organoids were grown in 3D Matrigel (356231, Corning) under prostate epithelial conditions⁴³. Cell viability was measured using 3D CellTiter-Glo 3D reagent (G9681, Promega) by quantifying metabolically active cells releasing ATP. Cell line-derived organoids were plated at a density of 2,000 cells per well in 96-well optical plates (3610, Corning) embedded in Matrigel as hanging drops (5 µl per well). Cells were treated with recombinant IL-23 (300-01A, PeproTech) at $100\, ng\, ml^{-1}$ or culture with ENZA ($10\, \mu M$) with or without recombinant IL-23. Luminescence measurements were performed after seven days in culture. Each IL-23 condition was normalized to its experimental control.

Characterization of the immune tumour microenvironment. Tumours were disaggregated and digested in collagenase D and DNase for 30 min at 37 °C to obtain a single-cell suspension. For intracellular cytokine detection, cells were stimulated for 5 h with PMA and ionomycin plus Golgi Plug. After neutralization of unspecific binding with a CD16/CD32 antibody (clone 93), single-cell suspensions were stained with specific monoclonal antibodies (primary antibodies directly conjugated) to assess the phenotype and diluted 1:200. The antibodies used were: CD45 (clone 30-F11, lot no. B235438); Ly-6G (clone 1A8, lot no. B194432); Ly6C (clone HK1.4, lot no. B243043), CD11b (clone M1/70, lot no. B233927); F4/80 (clone

BM8, lot no. 4305911), CD206 (clone C068C2, lot no. B230155), CD11c (clone N418, lot no. B226270), B220 (clone RA3-6B2, lot no. B210434), CD3 (clone 145-2C11, lot no. B241616), CD8 (clone 53-6.7, lot no. B193838), CD4 (clone GK1.5, lot no. B240053), NK1.1 (clone PK136, lot no. 4291566), CD90.2 (clone 30-H12, lot no. B190542), PDL1 (clone 10E.9G2, lot no. B191993), EpCAM (clone G8.8, lot no. B230070), pan-cytokeratin (clone C11, Lot. 4528S), IL-17 (clone TC11-18H10.1, lot no. B201753), IL-23p19 (clone FC23CPG, lot no. 4321359), isotype (rat IgG1, kappa, eBRG1) IL-23R (clone 12B2B64, lot no. 4321359). For flow gating, we used isotype controls of fluorescence minus one controls. All the antibodies were purchased from eBioscience or Biolegend. Samples were acquired on a BD Fortessa flow cytometer (BD Biosciences). Data were analysed using FlowJo software (TreeStar).

Immunohistochemistry and immunofluorescence of mouse tissues. For immunohistochemistry, tissues were fixed in 10% formalin (5701, ThermoScientific) and embedded in paraffin in accordance with standard procedures. Preceding immunohistochemical staining, tumour sections (4 µm) were exposed to two washes with OTTIX plus solution (X0076, Diapath) and subsequent hydration with OTTIX shaper solution (X0096, Diapath) followed by deionized water. Antigen unmasking was performed by heating sections in the respective pH solutions based on the antibodies used at 98 °C for 20 min. Subsequently the sections were blocked for peroxidases and nonspecific binding of antibodies using 3% H₂O₂ (23615.248, VWR) and Protein-Block solution (X0909, DAKO Agilent technologies), respectively, for 10 min each, split by 0.5% PBST washing. Haematoxylin and eosin staining was performed according to standard procedures. Sections were stained for anti-Ki-67 (clone SP6; Laboratory Vision Corporation), anti-pSTAT3 (Tyr705; clone D3A7; Cell Signaling). Images were obtained using objectives of 5×, 10× and $40 \times$ magnification and pixel images of $2.24 \mu m$, $1.12 \mu m$ and $0.28 \mu m$, respectively. All the quantifications were done using the public online software ImmunoRatio (http://153.1.200.58:8080/immunoratio/). For the immunofluorescence staining, tissue paraffin-embedded sections were stained with 4',6-diamidine-2'-phenylindole dihydrochloride (DAPI) (70238421, Roche), anti-IL-23 (ab45420, Abcam), anti-Ly6G (RB6-8C5, GeneTex). Confocal images were obtained with the Leica TCS SP5 confocal microscope using a $10 \times /1.25$ NA oil objective.

In vitro T cell suppression assay. In vitro suppression assays were carried out in RPMI with 10% FCS in 96-well U-bottom plates (Corning). Naive splenocytes were labelled with $5\,\mu\text{M}$ CFSE (Molecular Probes) and activated in vitro with anti-CD3 and anti-CD28 beads (Invitrogen) according to the manufacturer's instructions. Conditioned medium from BM-MDSCs was added to the culture. After three days, the proliferation of CFSE-labelled CD8⁺ T cells was analysed by BD Fortessa. **CRISPR-Cas9 transfection.** TRAMP-C1 cells were grown in 75-cm² flask to 50--60% confluency in DMEM medium supplemented with 10% heat-inactivated FBS, 100 U ml⁻¹ penicillin, 0.1 mg ml⁻¹ streptomycin and 2 mM L-glutamine. The transfection of the *Il23r* CRISPR-Cas9 KO plasmid (Santa Cruz Biotechnology) was performed using jetPRIME transfection reagent according to the manufacturer's protocol at the 1:2 DNA:jetPRIME ratio. After 24h of transfection, GFPtransduced cells were sorted to 99% purity and single cells were plated in 96-well plates. At day 7 after sorting, the grown cell clones were moved into 24-well plates for further expansion. The knockdown of the *Il23r* gene in each cell colony was confirmed by western blot.

NanoString. The nCounter analysis system (NanoString Technologies) was used to screen for the expression of signature genes associated with cancer–inflammation pathways. Two specific probes (capture and reporter) for each gene of interest were used. In brief, $5\,\mu l$ of RNA (the concentration is higher than $25\,ng\,\mu l^{-1}$) was hybridized with customized Reporter CodeSet and Capture ProbeSet as Mouse PanCancer Immune Profiling Panel including 700 selected genes (NanoString Technologies) according to the manufacturer's instructions for direct labelling of mRNAs of interest with molecular barcodes without the use of reverse transcription or amplification. Total RNA was quantified by NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilming-ton, DE) and RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). The hybridized samples were then recovered in the NanoString Prep Station and the mRNA molecules were counted with the NanoString nCounter. For expression analysis, each sample profile was normalized to geometric mean of 20 housekeeping genes included in the panel.

Immune tumour microenvironment characterization of tumours from patients with prostate cancer. Tumours were disaggregated and digested in collagenase I and DNase for 20 min at 37 °C to obtain single-cell suspensions. For intracellular cytokine detection, cells were stimulated for 5 h with PMA and ionomycin plus Golgi Plug. Single-cell suspensions were stained with specific monoclonal antibodies diluted 1:200 (primary antibodies directly conjugated) to assess the phenotype. The antibodies used were: CD45RA (clone MEM-56, 1:50), CD33 (clone WM53), CD11b (clone ICRF44), CD15 (clone W6D3), HLA-DR (clone L243), IL-23p19 (clone 23DCDP). For flow gating, we used isotype controls of fluorescence minus one controls. All antibodies were purchased from eBioscience or Biolegend.

Samples were acquired on a BD Fortessa flow cytometer (BD Biosciences). Data were analysed using FlowJo software (TreeStar.

Protein profiling. Plasma pools of patients with CSPC or CRPC were processed as indicated in the Human XL Cytokine Array Kit (R&D Systems). Pools of tissue lysates of tumours from sham and castrated $Pten^{PC-/-}$ mice were processed as indicated in the Mouse XL Cytokine Array Kit (R&D Systems). Developed films were scanned, the obtained images were analysed using ImageJ version 1.43u and background signals were subtracted from the experimental values.

Multiplex immunofluorescence in formalin-fixed paraffin-embedded tissue section. PMN-MDSC panel (CD15, CD11b, CD33 and EpCAM). Multiplex immunofluorescence for CD15 (M3631, Dako, clone Carb-3), CD33 (ab11032, Abcam, clone 6C5/2), CD11b (ab52477, Abcam, clone EP1345Y) and EpCAM Alexa Fluor 647 conjugate (5447S, Cell Signaling, clone VU1D9) was performed using 4-µm sections of formalin-fixed paraffin-embedded (FFPE) prostate tumour samples by sequential staining after antigen retrieval in CC1 (pH 8.5) (950-224, Ventana) in a water bath at 98 °C for 36 min. First, mouse monoclonal (IgG1) antibody anti-CD33 (1:100 dilution), mouse monoclonal (IgM) anti-CD15 (1:200 dilution) and rabbit monoclonal (IgG) antibody anti-CD11b (1:100 dilution) were incubated for 1 h after blocking with 10% goat serum for 30 min. Slides were then incubated with goat anti-mouse IgG1 Alexa Fluor 555-conjugated (A21127, Life Technologies), goat anti-mouse IgM Alexa Fluor 488-conjugated (A21042, Life Technologies) and goat anti-rabbit IgG (H+L) Alexa Fluor 700-conjugated (A21038, Life Technologies) antibodies for 30 min. Next, tissue sections were treated with 5% mouse or rabbit normal serum for 30 min, followed by incubation with a mouse monoclonal (IgG1) anti-EpCAM antibody conjugated to Alexa Fluor 647 (dilution, 1:200) for 1 h. The samples were washed three times for 5 min with TBS Tween 0.05% between incubations. Nuclei were counterstained with DAPI (70238421, Roche) and tissue sections were mounted with ProLong Gold antifade reagent (P36930, Molecular Probes).

CD15, IL-23 and EpCAM. Immunofluorescence was performed on 4-μm FFPE tissue sections using an automated staining platform (Bond-RX, Leica Microsystems). In brief, antigen retrieval was achieved using ER1 (pH 6.0) (AR9961, Leica Biosystems) for 30 min. Sections were blocked in 10% normal goat serum for 30 min at room temperature. Primary antibodies, including mouse monoclonal (IgM) anti-CD15 (M3631, Dako, clone Carb-3, dilution 1:200), rabbit monoclonal (IgG) anti-IL-23 (ab190356, Abcam, clone EPR5585(N), dilution 1:100) and mouse monoclonal (IgG1) anti-EpCAM (2929S, Cell Signaling, clone VU1D9, dilution 1:500), were incubated for 1 h. Slides were then incubated with goat anti-rabbit (H+L) Alexa Fluor 555-conjugated (A21429, Life Technologies), goat anti-mouse IgG1 Alexa Fluor 488-conjugated (A21042, Life Technologies) and goat anti-mouse IgG1 Alexa Fluor 647-conjugated (A21240, Life Technologies) antibodies for 30 min. Nuclei were counterstained with DAPI (70238421, Roche) and tissue sections were mounted with ProLong Gold antifade reagent (P36930, Molecular Probes).

Microscopy and image acquisition. After staining, slides were scanned using the multi-spectral camera provided by Vectra (Perkin Elmer) system. The number of images collected per case was dependent on tumour size from minimum of 1 to a maximum of 18 (average = 12). Quantification of PMN-MDSC-like cells (CD15⁺ CD33+CD11b+) was performed using inForm v.2.1.1 software (PerkinElmer) and the density of cells of interest are presented as the number of cells per mm². A tissue segmentation algorithm based on EpCAM positivity was used to separate tumour from adjacent stroma. The algorithm was trained to perform cell segmentation using counterstaining-based segmentation achieved with nuclear DAPI staining. Phenotype determination was based on positivity for CD15, CD33 and CD11b. Cells in tumour areas selected by the algorithm were then separated into bins as follows: CD15+CD33+CD11b+ cells were called PMN-MDSC-like cells and CD15-CD11b+ cells were called CD15-CD11b+ cells. All tissue segmentation, cell segmentation and phenotype determination maps were reviewed by a pathologist. Validation of antibody specificity for multiplex immunofluorescence. Immunohistochemistry was performed on 4- μm FFPE tissue sections using an automated staining platform (Bond-RX, Leica Microsystems). Optimal antibody concentrations were determined for primary antibodies against CD15 (M3631, Dako, clone Carb-3, dilution 1:200), CD33 (ab11032, Abcam, clone 6C5/2, dilution 1:100), CD11b (ab52477, Abcam, clone EP1345Y, dilution 1:100) IL-23 (ab190356, Abcam, clone EPR5585(N), dilution 1:100) and EpCAM (2929S, Cell Signaling, clone VU1D9, dilution 1:500). Antibody labelling was detected with the Bond Polymer Refine Detection Kit (DS9800, Leica Microsystems). 3,3-diaminobenzidine tetrahydroxychloride (DAB) was used as chromogen and the slides were counterstained with haematoxylin. Human control samples included colorectal specimens. In each staining batch, positive and negative controls were incubated with and without primary antibody.

RNA expression and qPCR. RNA isolation (TRIzol, Qiagen) and retro-transcription with SuperScriptIII (Invitrogen, 11752-250) were performed according to the manufacturer's instructions. qPCR reactions (Bio-Rad) were performed using

KAPA SYBR FAST qPCR green (KK4605; Applied Biosystems) and the specific primers reported below. Primer sequences were obtained from PrimerBank (http:// pga.mgh.harvard.edu/primerbank/index.html) or Bio-Rad. Each expression value was normalized to HPRT or GADPH level as reference. The primer sequences used were as follows: CXCL1 forward, 5'-CTGGGATTCACCTCAAGAACATC-3'; reverse, 5'-CAGGGTCAAGGCAAGCCTC-3'. CXCL2 forward, 5'-GCGTCACAC TCAAGCTCTG-3'; reverse, 5'-CCAACCACCAGGCTACAGG-3'; CXCL3 forward, 5'-ATCCCCCATGGTTCAGAAA-3'; reverse, 5'-ACCCTGCAGGAAG TGTCAAT-3'; CXCL5 forward, 5'-GTTCCATCTCGCCATTCATGC-3'; reverse, 5'-GCGGCTATGACTGAGGAAGG-3'. GAPDH forward, 5'-AGGT CGGTGTGAACGGATT-3'; reverse, 5'-TGTAGACCATGTAGTTGAG-3'. IL23p19 forward, 5'-CCAGCAGCTCTCTCGGAATC-3'; reverse, 5'-TCATATG TCCCGCTGGTGC-3'. Bio-Rad primers used were: Hprt PrimePCR PreAmp for SYBR Green Assay: Hprt, mouse qMmuCID0005679; Ar PrimePCR PreAmp for SYBR Green Assay: Ar, mouse qMmuCID0005164; Nkx3-1 PrimePCR PreAmp for SYBR Green Assay: Nkx3-1, mouse qMmuCED0046482; Pbsn PrimePCR PreAmp for SYBR Green Assay: Pbsn, mouse qMmuCID0017831; Fkbp5 PrimePCR PreAmp for SYBR Green Assay: Fkbp5, msouse qMmuCID0023283. Western blot analyses and protein detection. Tissue and cell lysates were prepared with RIPA buffer (1× PBS, 1% Nonidet P40, 0.5% sodium deoxycholate, 0.1% SDS and protease inhibitor cocktail; Roche). The total protein concentration was measured using a BCA Protein Assay Kit (23225; Pierce). Equal amounts of proteins were separated by SDS-PAGE and western blotted onto a 0.45-μm nitrocellulose membrane. Membranes were blocked in 5% defatted milk or 5% BSA in Tris-buffered saline containing 0.1% Tween-20 (TBST), probed with diluted antibodies and incubated at 4°C overnight. The following primary antibodies were used: rabbit polyclonal anti-HSP90 (1:1,000 dilution, Cell Signaling), rabbit polyclonal anti-pSTAT3 (Tyr705) (1:1,000 dilution, Cell Signaling), rat monoclonal anti-ROR7t (5:1,000 dilution, clone AFKJS-9, eBioscence), rabbit polyclonal anti-IL-23R (H-300) (1:1,000 dilution, Santa Cruz). After washing in TBST, the membrane was incubated with secondary antibodies that were conjugated to horseradish peroxidase (HRP) (1:5,000 dilution, Cell Signaling). The protein bands were visualized using the ECL Western Blotting Substrate (Pierce).

Samples from human prostates. Samples were acquired from patients with mCRPC, who had given their written informed consent to institutional protocols approved either by the Royal Marsden NHS Foundation Trust Hospital (London, UK) Ethics Committee (reference no. 04/Q0801/60), the IRCCS Ospedale San Raffaele (Milan, Italy) Ethics Committee (reference no. 99/INT/2004; 58/INT/2010) or the Azienda Ospedaliera di Padova (Padova, Italy) Ethics Committee (reference no. CESC/958P/2005). Human biological samples were sourced ethically and their research use was in accord with the terms of the informed consent provided. Fifty-one patients with CRPC treated at The Royal Marsden NHS Foundation Trust Hospital with sufficient formalin-fixed, paraffin-embedded, had matching CSPC and CRPC biopsies identified for multiplex immunofluorescence (see Supplementary Table 1). Four patients with CRPC, enrolled at Azienda Ospedaliera di Padova, and four patients with CSPC, enrolled at IRCCS Ospedale San Raffaele, were selected to perform the immune tumour microenvironment characterization by flow cytometry analyses. Case selection was blinded to baseline

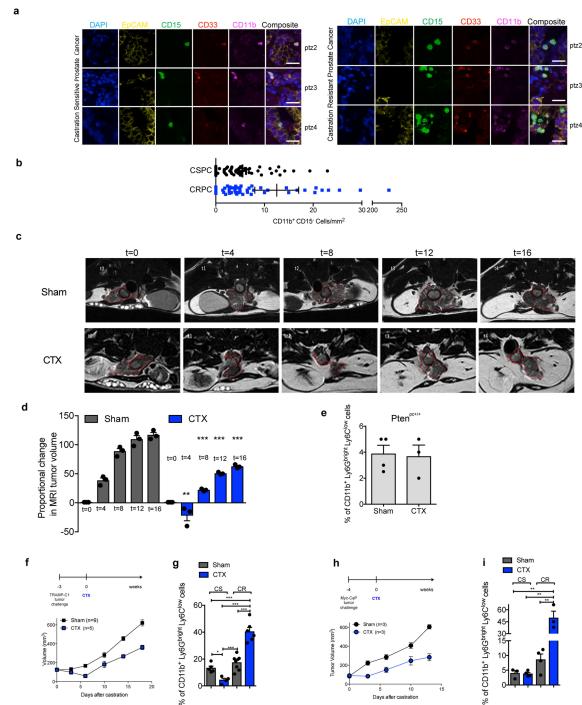
characteristics, treatments received, clinical outcome and molecular characterization to reduce any potential selection bias. Finally, plasma from 120 patients with CRPC with sufficient samples stored (including 28 plasma samples within 40 days of CRPC biopsy) and 20 patients with CSPC were analysed for IL-23 levels.

Statistical analysis and reproducibility. Data analyses were carried out using GraphPad Prism version 7. The data are mean ± s.e.m., individual values as scatter plots with column bar graphs and were analysed using Student's t-tests (paired or unpaired according to the experimental setting) by a two-sided test, and, when indicated, followed by Wilcoxon signed-rank test. One-way ANOVA was used to compare three or more groups in time point analyses. Differences were considered significant when P < 0.05 and are indicated as not significant, *P < 0.05, **P < 0.01, ***P < 0.001. Non-parametric tests were applied when variables were not normally distributed using the SPSS statistical software. n values represent biological replicates. Survival curves were compared using the log-rank test (Mantel-Cox). Because of evidence of overdispersion, tumour-infiltrating PMN-MDSC (CD15⁺CD11b⁺CD33⁺EpCAM⁻) and CD15⁻CD11b⁺ cells were analysed using mixed-effect negative binomial regression model (with per patient random intercept) when comparing paired biopsies, and a negative binomial regression model was used when analysing the association between CRPC biopsies and IL-23. PMN-MDSC (coefficient, 1.49; 95% confidence interval, 0.83-2.15; P < 0.001); CD15⁻CD11b⁺ cells (coefficient, 0.43; 95% confidence interval, 0.04–0.83; not significant (P > 0.05)). All statistics and reproducibility information are reported in the figure legends. For animal studies, sample size was defined on the basis of past experience with the models¹⁵, to detect differences of 20% or greater between the groups (10% significance level and 80% power). For ethical reasons the minimum number of animals necessary to achieve the scientific objectives was used. Animals were allocated randomly to each treatment group. Different treatment groups were processed identically and animals in different treatment groups were exposed to the same environment.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Source Data for the figures and extended data figures are provided in the online version of the paper. CSPC and mCRPC tumour biopsy mRNA-seq data that support the findings of this study are available in the SU2C-PCF IDT cBioportal (http://www.cbioportal.org) and through dbGAP (https://www.ncbi.nlm.nih.gov/gap/) with the identifier phs000915.v1.p1⁴⁰.

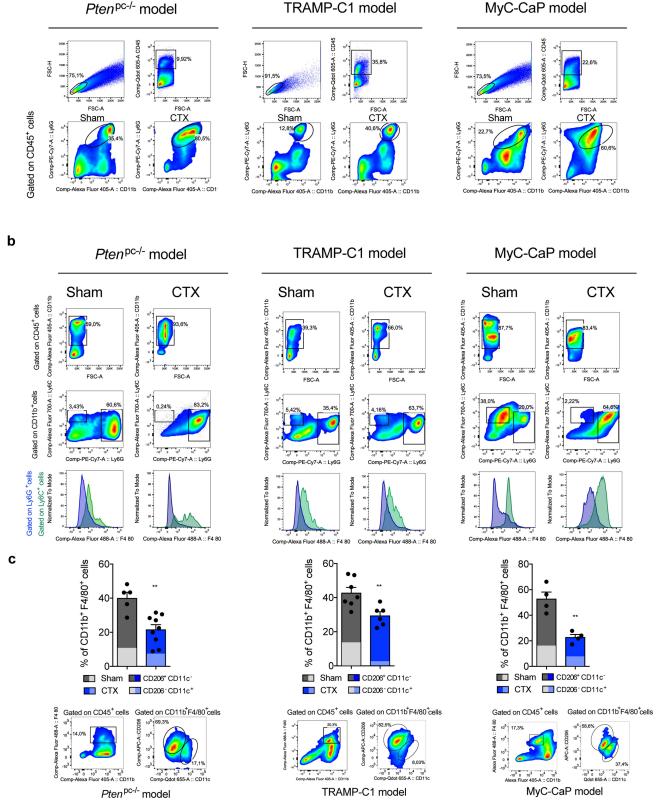
- Marigo, I. et al. Tumor-induced tolerance and immune suppression depend on the C/EBPβ transcription factor. Immunity 32, 790–802 (2010).
- Lechner, M. G., Liebertz, D. J. & Epstein, A. L. Characterization of cytokineinduced myeloid-derived suppressor cells from normal human peripheral blood mononuclear cells. *J. Immunol.* 185, 2273–2284 (2010).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215–1228 (2015).
- Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36 (2013).
- Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578 (2012).
- Drost, J. et al. Organoid culture systems for prostate epithelial and cancer tissue. Nat. Protoc. 11, 347–358 (2016).



Extended Data Fig. 1 \mid Multispectral images of PMN-MDSCs in human biopsies and set-up of the different CRPC mouse models. a, Multispectral microscopy images (EpCAM, yellow; CD15, green; CD33, red; CD11b, pink) of castration-sensitive and castration-resistant prostate cancers. n = 3 biological independent patients. Scale bars, 20 μm. **b**, Quantification of the number of CD11b⁺CD15⁻ cells in the tumour of castration-sensitive and castration-resistant prostate cancers (CSPC, n = 51; CRPC, n = 51 biological independent patients). Cells were counterstained with the nuclear marker DAPI (blue). Statistical analyses (negative binomial regression model): P = 0.062. c, MRIs of one representative sham-operated (Sham) or surgically castrated (CTX) $Pten^{PC-/-}$ mouse of the three analysed at different time points. **d**, Waterfall plot depicting proportional change in tumour response for sham (n = 3)and CTX (n=3) Pten^{PC-/-} mice. **e**, Prostate PMN-MDSC frequencies determined by flow cytometry in sham (n=3) and CTX (n=3) Pten^{pc+/+} mice (12 weeks after castration). Statistical analyses (two-sided unpaired Student's *t*-test): P = 0.85. **f**, Schematic representation of the experiment.

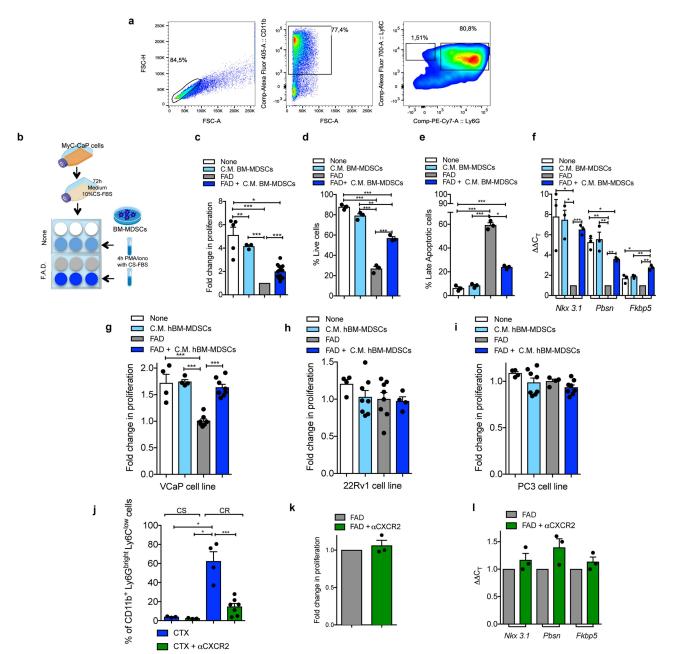
Six-week-old C57BL/6 males were challenged subcutaneously with TRAMP-C1 cells. When tumours reached approximately 100 mm³, mice were sham-operated (sham, n = 9) or surgically castrated (CTX, n = 5). g, Tumour PMN-MDSC frequencies were determined by flow cytometry during castration-sensitive and castration-resistant phases. Sham CSPC, n = 5; CTX CSPC, n = 4; sham CRPC, n = 8; CTX CRPC n = 6. h, Schematic representation of the experiment. Six-week-old FVB males were challenged subcutaneously with MyC-CaP cells. When tumours reached approximately 100 mm^3 , mice were sham-operated (sham, n = 3) or surgically castrated (CTX, n = 3). i, Tumour PMN-MDSC frequencies were determined by flow cytometry during castration-sensitive and castration-resistant phases. Sham CSPC, n = 3; CTX CSPC, n = 4; sham CRPC, n = 4; CTX CRPC, n = 3. **b**, **d**, **e**, **g**, **i**, Data are mean \pm s.e.m. d, g, i, Statistical analyses (unpaired two-sided Student's t-test): ns, not significant; *P < 0.05; **P < 0.01; ***P < 0.001. **f**, **h**, Statistical analyses (two-sided unpaired Student's t-test followed by Wilcoxon signed-rank test): *P < 0.05.

а



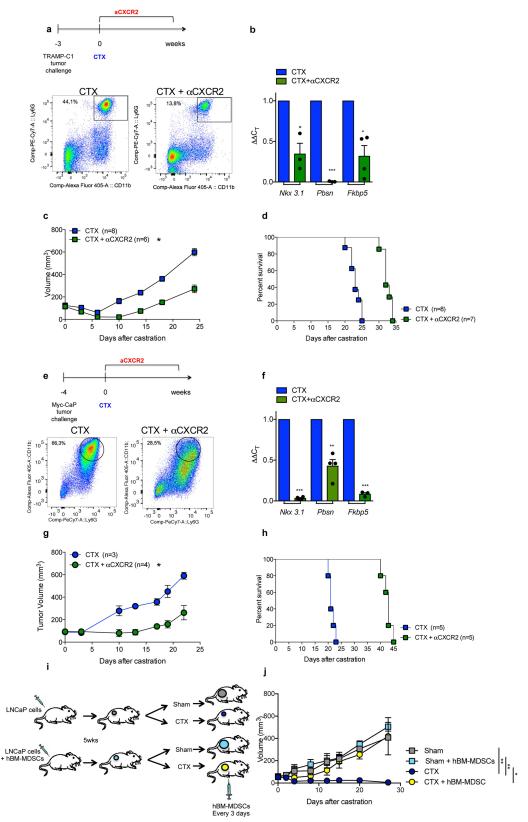
Extended Data Fig. 2 | Multiparametric flow cytometry analysis of myeloid populations in the tumours of different CRPC mouse models. a, Gating strategy relative to the quantification of PMN-MDSCs in the tumour of sham- and CTX-operated mice for $Pten^{PC-/-}$, TRAMP-C1 and MyC-CaP models. b, Representative dot plot reporting the Ly6G⁺ and Ly6C⁺ cells in the tumour of sham- and CTX-operated mice for $Pten^{PC-/-}$, TRAMP-C1 and MyC-CaP models. Data were validated in at

least three biological independent mice. c, Representative dot plot and quantification of CD11b+F4/80+ cells and CD11c+CD206- (M1-like) or CD11c-CD206+ (M2-like) macrophages in the tumour of sham- and CTX-operated mice for $Pten^{PC-/-}$ (sham, n=5; CTX, n=9), TRAMP-C1 (sham, n=7; CTX, n=6) and MyC-CaP (sham, n=4; CTX, n=4) models. Data are mean \pm s.e.m. Statistical analyses (unpaired two-sided Student's t-test): **P < 0.01.



Extended Data Fig. 3 | Factors secreted by MDSCs promote insensitivity to ADT in androgen-dependent mouse and human prostate cancer cell lines and the CXCR2 antagonist impairs tumour recruitment of MDSCs in *Pten*^{PC-/-} mice. a, Representative dot plot reporting the BM-MDSCs after in vitro differentiation. Data were validated in two biological independent experiments. **b**, Experimental scheme. **c**, Cell proliferation of MyC-CaP cells (none, n = 5; conditioned medium (C.M.) from BM-MDSCs, n = 3; FAD, n = 13; FAD and BM-MDSCs, n = 13 biological independent samples). **d**, Percentage of annexin V⁻ and 7AAD⁻ MyC-CaP cells. **e**, Percentage of annexin V⁺ and 7AAD⁺ MyC-CaP cells. **f**, qRT-PCR analyses of the indicated genes in MyC-CaP cells. **g**, **h**, **i**, Cell proliferation of VCaP (none, n = 4; conditioned medium from human BM-MDSCs, n = 4; FAD, n = 7; FAD and human BM-MDSCs, n = 8 biological independent samples), 22Rv1 (none, n = 4; conditioned medium from human BM-MDSCs, n = 8; FAD, n = 8; FAD and human BM-MDSCs, n = 4 biological independent samples) and PC3 (none, n = 4; conditioned medium from human BM-MDSCs, n = 8; FAD, n = 4; FAD

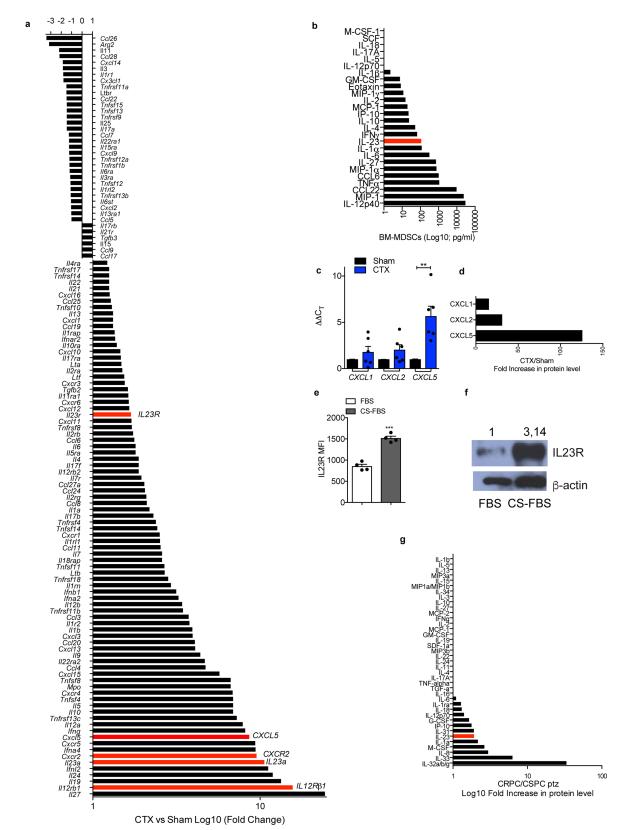
and human BM-MDSCs, n = 8 biological independent samples) prostate cancer cells. j, Tumour MDSC frequencies determined by flow cytometry of prostate tumours of CTX Pten^{PĈ-/-} mice treated or not with CXCR2 antagonist (α CXCR2) at completion of the study (12 weeks after CTX). CTX castration-sensitive, n = 3; CTX and CXCR2 antagonist castrationsensitive, n = 3; CTX castration-resistant, n = 4; CTX and CXCR2 antagonist castration-resistant, n = 7 biological independent mice. **k**, Cell proliferation of TRAMP-C1 cells after 72 h of treatment with CXCR2 antagonist. I, qRT-PCR analyses of the indicated genes in TRAMP-C1 cells after 24 h of treatment (fold change compared to the FAD condition). k, l, Aggregated data from three independent experiments are reported, fold change compared to the FAD condition. c-l, Data are mean \pm s.e.m. \mathbf{c} - \mathbf{f} , n = 3 biological independent samples. \mathbf{d} , \mathbf{e} , \mathbf{g} - \mathbf{j} , Statistical analyses (unpaired two-sided Student's *t*-test): ns, not significant; *P < 0.05; **P < 0.01; ***P < 0.001. **c**, **f**, Statistical analyses (two-sided unpaired Student's *t*-test followed by Wilcoxon signed-rank test): *P < 0.05.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Impaired tumour recruitment of MDSCs enhances response to surgical castration in different allograft models of prostate cancers. a, Schematic representation of the experiment. Six-week-old C57BL/6 males were challenged subcutaneously with TRAMP-C1 cells. When tumours reached approximately 100 mm³, mice were surgically castrated and left untreated (CTX, n = 8) or treated with CXCR2 antagonist (CTX and CXCR2 antagonist, n = 9). Representative flow cytometry plots of PMN-MDSCs (CD11b+Ly6G+ cells, gated on CD45⁺ cells) in tumours for each experimental condition. **b**, qRT–PCR analyses of the indicated genes in the prostate tumours after CTX or CTX and CXCR2 antagonist treatment (n = 3 per group). Data are mean \pm sem. Statistical analyses (unpaired two-sided Student's *t*-test): **P* < 0.05; ***P < 0.001. c, Mean tumour volume (\pm s.e.m.) for each experimental group. Statistical analyses (unpaired two-sided Student's t-test followed by Wilcoxon signed-rank test): ***P < 0.001. **d**, Survival curves are reported in Kaplan-Meier plot. Statistical analyses (two-sided log-rank test): ***P < 0.001. **e**, Schematic representation of the experiment. Six-week-old FVB males were challenged subcutaneously with MyC-CaP cells. When tumours reached approximately 100 mm³, mice were surgically castrated and left untreated (CTX, n = 5) or treated with CXCR2 antagonist (CTX

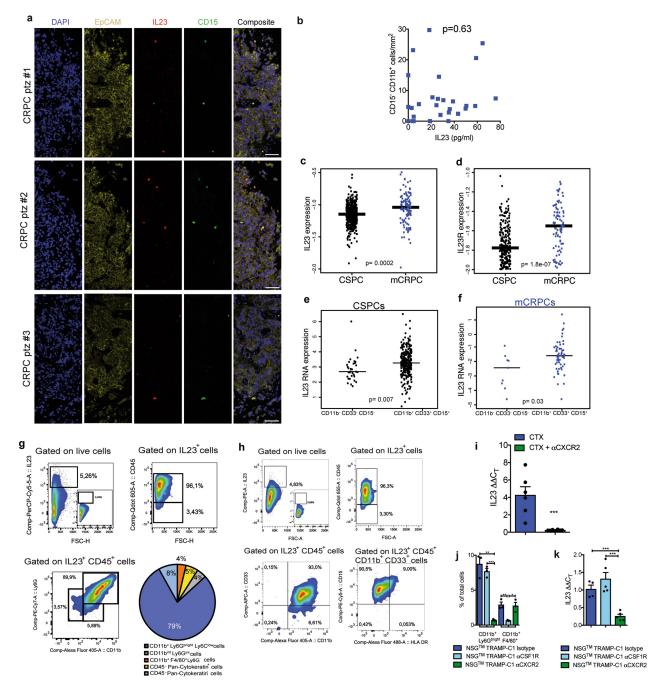
and CXCR2 antagonist, n = 5). Representative flow cytometry plots of PMN-MDSCs (CD11b⁺Ly6G⁺ cells, gated on CD45⁺ cells) in tumours for each experimental condition. f, qRT-PCR analyses of the indicated genes in the prostate tumours after CTX or CTX and CXCR2 antagonist treatment (n = 3 per group). Data are mean \pm s.e.m. Statistical analyses (unpaired two-sided Student's t-test): **P < 0.01; ***P < 0.001. g, Average tumour volume (\pm s.e.m.) for each experimental group. Statistical analyses (two-sided unpaired Student's t-test followed by Wilcoxon signed-rank test): *P < 0.05. **h**, Survival curves reported as Kaplan–Meier plot. Statistical analyses (two-sided log-rank test): **P< 0.01. **i**, Schematic representation of the experiment. Six-week-old NOD/SCID males were challenged subcutaneously with LNCaP cells or with LNCaP cells and human BM-MDSCs. When tumours reached approximately 70 mm³, mice were sham-operated (sham, n = 5) or sham-operated and injected every three days intraperitoneally with 3 \times 10⁶ human BM-MDSCs (sham and human BM-MDSCs, n = 5) or surgically castrated and left untreated (CTX, n = 8) or treated with human BM-MDSCs (CTX and human BM-MDSCs, n = 5). **j**, Average tumour volume (\pm s.e.m.) for each experimental group. Statistical analyses (unpaired two-sided Student's t-test followed by Wilcoxon signed-rank test): **P < 0.01.



Extended Data Fig. 5 | See next page for caption.

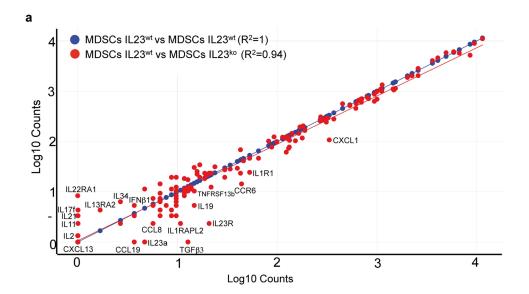
Extended Data Fig. 5 | IL-23 pathway is the most upregulated in the tumour after castration. a, Gene expression of selected genes determined by NanoString nCounter gene expression assay in sham $Pten^{PC-/-}$ and CTX $Pten^{PC-/-}$ tumours. Data are shown as pool of n=5. b, Analyses of the conditioned medium of bone marrow-derived MDSCs tested for the indicated soluble molecules by Mouse CytokineMAP B version 1.0. The graph shows the concentration of the indicated soluble molecules as \log_{10} of the concentration found in the conditioned medium of BM-MDSCs, the values were subtracted of the background (culture medium). Data are shown as pool of n=10. c, qRT-PCR analyses of the indicated genes in sham (n=6) and CTX (n=6) $Pten^{PC-/-}$ tumours. Data are mean \pm s.e.m. of biological independent animals. Statistical analyses (unpaired two-sided

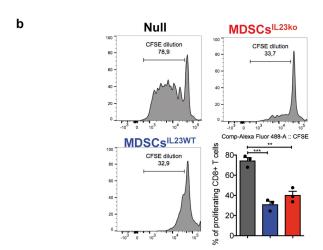
Student's *t*-test): *P<0.05. **d**, Protein level of CXCL1, CXCL2 and CXCL5 in CTX $Pten^{PC-/-}$ tumours. Data are analysed as ratio between CTX (pool of three samples) and sham (pool of three samples) $Pten^{PC-/-}$ tumours and reported as fold increase in protein level. **e**, **f**, IL-23R protein level analysed by flow cytometry and western blot on TRAMP-C1 cells under normal culture conditions (FBS) or androgen-deprivation culture conditions (charcoal-stripped FBS). n = 4 biological independent samples per group. **f**, Numbers indicate fold change in protein level. Loading control: anti- β -actin antibody. The western blot was validated at least twice. **g**, Protein profile of the plasma of patients with CSPC and CRPC. Data are analysed as ratio between CRPC (pool of 18 samples) and CSPC (pool of 17 samples) and reported as fold increase in protein level.



Extended Data Fig. 6 | Characterization of IL-23+ cells in the tumour of CTX $Pten^{PC-/-}$ mice and patients with CRPC. a, Multispectral microscopy images (EpCAM, yellow; CD15, green; IL-23, red) of three patients with CRPC. b, Correlation analyses of the numbers of CD15-CD11b+ cells in the tumour and IL-23 levels in the plasma of patients of CRPC (n=28). Statistical analyses (negative binomial regression model): P=0.63. c, d, IL23A and IL23R mRNA expression in the tumour of CSPCs (n=549) and mCRPCs (n=116). e, f, Expression of IL-23 in PMN-MDSC marker-positive (CD11b+CD33+CD15+) tumours from patients with CSPC or mCRPC. c-f, Statistical analyses (unpaired two-sided Student's t-test) are reported. g, Representative plots of IL-23+, CD45+ and CD45-, Ly6GbrightCD11b+ and Ly6GintCD11bint, CD11b+F4/80+ cells pregated on the reported population in the tumour of CTX $Pten^{PC-/-}$ mice. IL-23 gate was decided based on isotype control

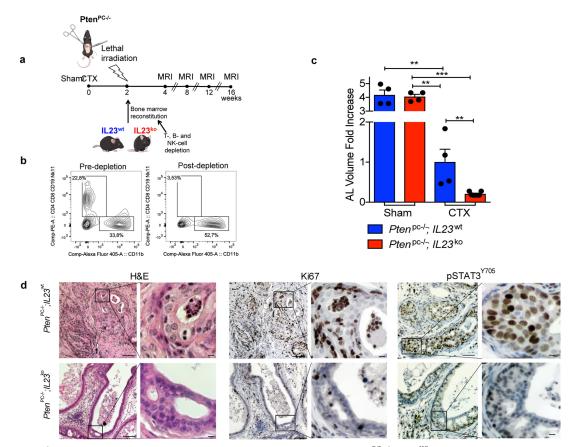
panel (insert). Pie chart shows the percentage of the different subsets gated on IL-23+ cells in the tumour of $Pten^{PC-/-}$ mice (mean, n=9). \mathbf{h} , qRT–PCR analyses of IL-23 in the prostate tumours of castrated (CTX; n=6) or castrated and treated with CXCR2 antagonist (CTX+CXCR2 antagonist; n=7) $Pten^{PC-/-}$ mice. Data are mean \pm s.e.m. \mathbf{i} , PMN-MDSC and TAM frequencies determined by flow cytometry in the tumour of castrated NSG TRAMP-C1 allografts upon treatment with isotype, CSF1R antibody, CXCR2 antagonist. Data are mean \pm s.e.m. (n=3 per group). \mathbf{j} , qRT-PCR analyses of IL-23 in the tumour of castrated NSG TRAMP-C1 allografts upon treatment with isotype (n=4), CSF1R antibody (n=5), CXCR2 antagonist (n=5). Data are mean \pm s.e.m. Each dot represents a biological independent animal. \mathbf{h} - \mathbf{j} , Statistical analyses (unpaired two-sided Student's t-test): *P < 0.05; **P < 0.01; ***P < 0.01.





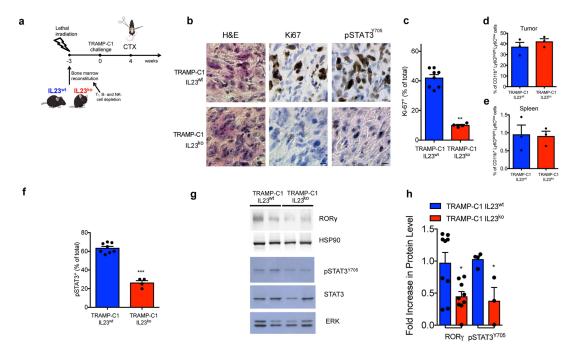
Extended Data Fig. 7 | $Il23a^{WT}$ and $Il23a^{KO}$ BM-MDSCs have equal secretome and equal immunosuppressive capability. a, Correlation analyses of the gene expression of selected genes determined by NanoString nCounter gene expression assay in $Il23a^{WT}$ and $Il23a^{KO}$ BM-MDSCs shown as scatter plot analyses. Data are shown as pool of n=10. R^2 values were determined using linear regression analyses.

b, Representative flow cytometry plots of CFSE dilution (gated on CD8 $^+$ cells) and quantification of proliferating CD8 $^+$ T cells exposed to conditioned medium from $Il23a^{WT}$ and $Il23a^{KO}$ BM-MDSCs. Data are mean \pm s.e.m. of three independent biological samples. Statistical analyses (unpaired two-sided Student's t-test): **P< 0.01; ****P< 0.01.



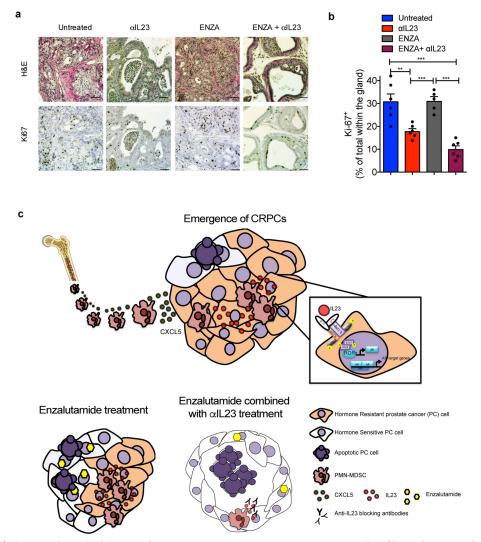
Extended Data Fig. 8 | Genetic inhibition of IL-23 limits resistance to castration in prostate cancer in $Pten^{PC-/-}$ mice. a, Experimental set-up. Sham-operated (sham) or castrated (CTX) $Pten^{PC-/-}$ mice were lethally irradiated and transplanted with bone marrow precursors depleted of T, B and NK cells from $Il23a^{WT}$ and $Il23a^{KO}$ mice. The mice were then monitored by MRI for tumour progression. b, Representative dot plot of bone marrow precursors pre- and post-depletion of T, B, and NK cells. Data were validated in two biological independent experiments. c, Quantification of the tumour size of sham-operated $Pten^{PC-/-}Il23a^{WT}$ (n=4) and $Pten^{PC-/-}Il23a^{KO}$ (n=4), and castrated $Pten^{PC-/-}Il23a^{WT}$

(n=4) and $Pten^{PC-/-}1l23a^{KO}$ (n=7) mice at the completion of the study is reported as fold increase of the prostate anterior lobe (AL) volume (fold change compared to the CTX $Pten^{PC-/-}1l23a^{WT}$ group). Data are mean ± s.e.m. Statistical analyses (unpaired two-sided Student's t-test): **P < 0.01, ****P < 0.001. **d**, Haematoxylin and eosin, Ki-67 and pSTAT3(Y705) immunohistochemical staining (Ki-67 and pSTAT3(Y705), brown; nuclei, blue) of one representative $Pten^{PC-/-}1l23a^{WT}$ and $Pten^{PC-/-}1l23a^{KO}$ mouse of at least three mice analysed at completion of the study. Scale bars, 50 μm (left) and 25 μm (right).



Extended Data Fig. 9 | Genetic inhibition of IL-23 limits resistance to castration in prostate cancer in TRAMP-C1 allograft model in vivo. a, Schematic representation of the experiment. Six-week-old C57BL/6 males were lethally irradiated and transplanted with bone marrow precursors from $Il23a^{\dot{W}T}$ and $Il23a^{KO}$ mice. After the bone marrow engraftment, the animals were challenged subcutaneously with TRAMP-C1 cells. When tumours reached approximately 100 mm³, mice were surgically castrated and monitored for tumour progression. b, Haematoxylin and eosin, Ki-67 and pSTAT3(Y705) immunohistochemical staining (Ki-67 and pSTAT3(Y705), brown; nuclei, blue) of representative TRAMP-C1 $Il23a^{WT}$ and TRAMP-C1 $Il23a^{KO}$ mice. Scale bars, 25 μ m. **c**, Quantification of Ki-67 $^+$ cells is reported as a percentage of the total number of cells. TRAMP-C1 $Il23a^{WT}$ (n=8) and TRAMP-C1 $Il23a^{KO}$ (n=4), one tumour per mouse, mean of three sections per tumour, ≥3 fields per section. Data are mean \pm s.e.m. of biologically independent mice. Statistical analyses (unpaired two-sided Student's t-test): **P < 0.01. **d**, **e**, PMN-MDSC

frequencies determined by flow cytometry in the tumour and in the spleen of TRAMP-C1 $Il23a^{WT}$ (n=3) and TRAMP-C1 $Il23a^{KO}$ (n=3) mice 10 days after castration. Data are mean \pm s.e.m. ${\bf f}$, Quantification of pSTAT3(Y705) reported as a percentage of the total number of cells. TRAMP-C1 $Il23a^{WT}$ (n=8) and TRAMP-C1 $Il23a^{KO}$ (n=4), one tumour per mouse, mean of three sections per tumour, ≥ 3 fields per section. Statistical analyses (unpaired two-sided Student's t-test): ***P < 0.001. ${\bf g}$, Western blot for ROR ${\bf \gamma}$, pSTAT3(Y705) and total STAT3 levels in prostate tumours of TRAMP-C1 $Il23a^{WT}$ and TRAMP-C1 $Il23a^{KO}$ mice. Loading control: HSP90 antibody or total ERK antibody. The western blot was validated at least twice. ${\bf h}$, Quantification is reported as mean \pm s.e.m. of biological independent experiments: TRAMP-C1 $Il23a^{WT}$ ROR ${\bf \gamma}$, n=9; and TRAMP-C1 $Il23a^{KO}$ ROR ${\bf \gamma}$, n=9; TRAMP-C1 $Il23a^{WT}$ ROR ${\bf \gamma}$. n=9; and TRAMP-C1 $Il23a^{KO}$ ROR ${\bf \gamma}$, n=9; TRAMP-C1 $Il23a^{WT}$ ROR ${\bf \gamma}$. n=9; TAT3(Y705), n=4; and TRAMP-C1 $Il23a^{KO}$ pSTAT3(Y705), n=3. Statistical analyses (unpaired two-sided Student's t-test): *P < 0.05.



Extended Data Fig. 10 | Pharmacological inhibition of IL-23 in association with ADT delays disease progression in prostate cancer. a, Haematoxylin and eosin and Ki-67 immunohistochemical staining (Ki-67, brown; nuclei, blue) of representative castrated $Pten^{PC-/-}$ mice treated with IL-23, ENZA or both. Scale bars, $50\,\mu\text{m}$. b, Quantification of Ki-67 reported as a percentage of total within the glands. One tumour per mouse, three sections per tumour, ≥ 3 fields per section. Data are mean \pm s.e.m. of biologically independent animals. Untreated, n=6; IL-23, n=6; ENZA, n=5 or both, n=6. Statistical analyses (unpaired two-sided Student's t-test): **P < 0.01; ***P < 0.001. c, After castration,

PMN-MDSCs progressively infiltrate the tumour bed mainly recruited by CXCL5. Within the tumour, PMN-MDSCs start to produce higher amount of IL-23, thus establishing a positive-feedback loop that induces the overexpression of IL-23R on the tumour epithelial cells and confer resistance to castration in prostate cancer by activating the STAT3–ROR γ pathway. ENZA treatment can block the AR, inducing sensitiveness of prostate cancer cells to androgen deprivation, but the persistent presence of PMN-MDSC-derived IL-23 rescues the drug sensitiveness leading to ADT resistance. Anti-IL-23 treatment reinstates sensitivity to castration in prostate cancer enhancing the efficacy of ENZA.



Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks

Kevin M. Cherry¹ & Lulu Qian^{1,2}*

From bacteria following simple chemical gradients¹ to the brain distinguishing complex odour information², the ability to recognize molecular patterns is essential for biological organisms. This type of information-processing function has been implemented using DNA-based neural networks³, but has been limited to the recognition of a set of no more than four patterns, each composed of four distinct DNA molecules. Winner-take-all computation⁴ has been suggested^{5,6} as a potential strategy for enhancing the capability of DNA-based neural networks. Compared to the linear-threshold circuits⁷ and Hopfield networks⁸ used previously³, winner-takeall circuits are computationally more powerful⁴, allow simpler molecular implementation and are not constrained by the number of patterns and their complexity, so both a large number of simple patterns and a small number of complex patterns can be recognized. Here we report a systematic implementation of winner-take-all neural networks based on DNA-strand-displacement^{9,10} reactions. We use a previously developed seesaw DNA gate motif^{3,11,12}, extended to include a simple and robust component that facilitates the cooperative hybridization¹³ that is involved in the process of selecting a 'winner'. We show that with this extended seesaw motif DNA-based neural networks can classify patterns into up to nine categories. Each of these patterns consists of 20 distinct DNA molecules chosen from the set of 100 that represents the 100 bits in 10×10 patterns, with the 20 DNA molecules selected tracing one of the handwritten digits '1' to '9'. The network successfully classified test patterns with up to 30 of the 100 bits flipped relative to the digit patterns 'remembered' during training, suggesting that molecular circuits can robustly accomplish the sophisticated task of classifying highly complex and noisy information on the basis of similarity to

Winner-take-all computation⁴ is one of the simplest competitive neural-network models, inspired by the lateral inhibition and competition observed among biological neurons in the brain¹⁴. In this model, the output of a neuron is ON if and only if the weighted sum of all binary inputs is the largest among all neurons (Fig. 1a). Here, in a winner-take-all neural network, the weight matrix associated with each output is referred to as a 'memory'. As shown in Fig. 1b, a simple training algorithm involves using the target patterns as weights. The example network has two memories—in other words, it 'remembers' two patterns—'L' and 'T'. The network 'recognizes' a pattern by comparing it to all memories and identifying which memory the pattern is most similar to—the output associated with this memory will be ON and all other outputs will be OFF. For instance, a corrupted 'L' with the last bit flipped from 1 to 0 can be recognized as 'L', because it will result in y_1 (the output of the neuron remembering 'L') being ON and y_2 (the output of the neuron remembering 'T') being OFF.

The winner-take-all function can be broken into five subfunctions, each of which can be implemented with a simple chemical reaction (Fig. 1c): First, weight multiplication of $x_i \times w_{ij}$ (where x_i is a binary input and w_{ij} is an analogue weight) is implemented with reactions wherein an input species X_i catalytically converts a weight species W_{ij} to an intermediate product P_{ij} . If X_i is absent, then no P_{ij} will be

produced; if X_i is present, then the final concentration of P_{ij} will be determined by the initial concentration of W_{ij} , thus setting the value of the weighted input. Second, summation is implemented with reactions that convert all intermediate species P_{ij} within the same neuron to a common weighted-sum species S_j . Third, comparison of weighted sums to determine which is the largest is implemented with a set of 'pairwise annihilation' reactions, wherein each weighted-sum species S_j destroys any other weighted-sum species S_k until only a single winner remains. Fourth, signal-restoration reactions bring the concentration of the winner species back to a predetermined output value—the final concentration of a winning output species Y_j corresponds to the initial concentration of a restoration-gate species RG_j . Last, reporting reactions are used to convert each output Y_i to a fluorescent signal Fluor_i.

All reactions except pairwise annihilation and signal restoration naturally take place sequentially, because the product of a previous reaction is a reactant of the next one. Because there are common reactants in the annihilation and restoration reactions, we used different rates to control their order: the former has a much faster rate constant than the latter, so a winner that survives all fast competitions is then converted slowly to an output signal.

Weight multiplication and signal restoration are both catalytic reactions, implemented with a pair of seesawing reactions 11 (Fig. 1e, Extended Data Fig. 1). An input X_i (or weighted sum S_j) species first interacts with a weight W_{ij} (or restoration gate RG_j) species through a reversible strand-displacement reaction 15 to release an intermediate product P_{ij} (or output Y_j) species. A fuel strand XF_i (or YF_j) then frees the input (or weighted sum) species for more catalytic cycles. As long as the fuel strand is in excess, all weight (or restoration gate) molecules will eventually be converted to intermediate (or output) molecules. Summation is implemented with a single seesawing reaction facilitated by a summation gate SG_j (Extended Data Fig. 1). The reaction is reversible by itself but drained forward by the downstream irreversible reaction of pairwise annihilation.

The annihilation reaction is implemented with cooperative hybridization¹³ (Fig. 1f). One weighted-sum strand S_i can bind to a toehold on one side of an annihilator molecule Anh_{ik} and branch-migrate to the middle point of the double-stranded domain. If only S_i is present, then this process is completely reversible and no molecules will be consumed. However, if another weighted-sum strand S_k is also present, then it can bind to another toehold on the opposite side of the annihilator and also branch-migrate to the middle point of the double-stranded domain. When the S_i and S_k strands reach the middle point simultaneously, the annihilator will be split apart into two waste molecules. Because neither waste molecule has a toehold exposed, it cannot interact with any other molecules. The annihilation reaction shown in Fig. 1f is designed to be roughly 100 times faster than the signal-restoration reaction shown in Fig. 1e, owing to the two extra nucleotides in both toeholds on the annihilator—it is known that the rate of strand displacement reactions grows exponentially faster with

Reporting is implemented with an irreversible strand-displacement reaction, wherein an output strand Y_i interacts with a double-stranded

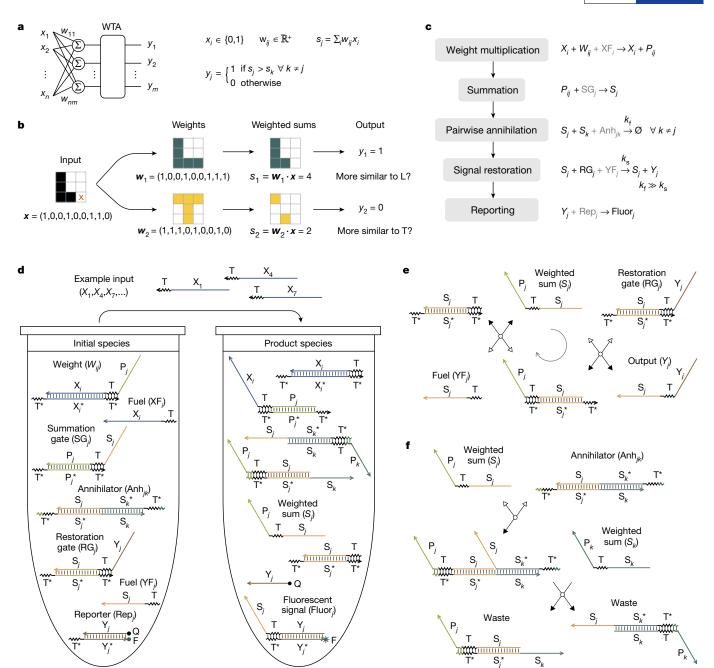


Fig. 1 | Winner-take-all neural network and its DNA implementation. **a**, A winner-take-all (WTA) neural network with *m* memories that each has *n* bits; x_1 to x_n and y_1 to y_m are binary inputs and outputs, respectively; w_{ij} $(1 \le i \le n \text{ and } 1 \le j \le m)$ are analogue weights of positive, real numbers; s_i and s_k $(1 \le j \ne k \le m)$ are weighted sums of the inputs. **b**, Example pattern recognition using target patterns as weights. Each 9-bit pattern is shown in a 3×3 grid. Each black or coloured pixel indicates a 1 and each white pixel indicates a 0. The two target patterns correspond to the letters 'L' and 'T', respectively. If the input pattern is corrupted (for example, the last bit of 'L' is flipped from 1 to 0, as indicated by the orange cross), then the neural network can still recognize it as being more similar to 'L' than to 'T', because the weighted sum using 'L' as weights is still larger than the weighted sum using 'T' as weights. c, Chemical-reaction-network implementation. The concentrations of chemical species X_i , W_{ij} , S_i and Y_i correspond to the values of variables x_i , w_{ij} , s_i and y_i , respectively. The species in black are needed as part of the function, whereas the species in grey are needed to facilitate the reactions. The waste molecules are not shown in the reactions. k_f and k_s are the rate constants of the pairwiseannihilation and signal-restoration reactions, respectively. d, DNA-strand-

displacement implementation. The initial test tube (left) shows all DNA species with $1 \le i \le n$ and $1 \le j \ne k \le m$. The final test tube (right) shows only the product species after a set of input strands are added, with i, j and k being a subset of all possible numbers depending on the specific input. Zigzag lines indicate short (5 or 7 nucleotide) toehold domains and straight lines indicate long (15 or 20 nucleotide) branch-migration domains in DNA strands, with arrowheads marking their 3' ends. Each domain is labelled with a name and assigned a unique DNA sequence, with asterisks in the names indicating sequence complementarity. Strand modifications are labelled as F and Q, where F indicates a fluorophore and Q indicates a quencher. e, Signal-restoration reaction. The grey circle with an arrow indicates the direction of the catalytic cycle. f, Pairwiseannihilation reaction. Representative (not all possible) states are shown. In e and f, arrows with black-filled and white-filled arrowheads indicate the forwards and backwards directions of a reaction step, respectively. The mechanisms of weight multiplication, summation and reporting reactions are shown in Extended Data Fig. 1. DNA sequences are listed in Supplementary Table 1.

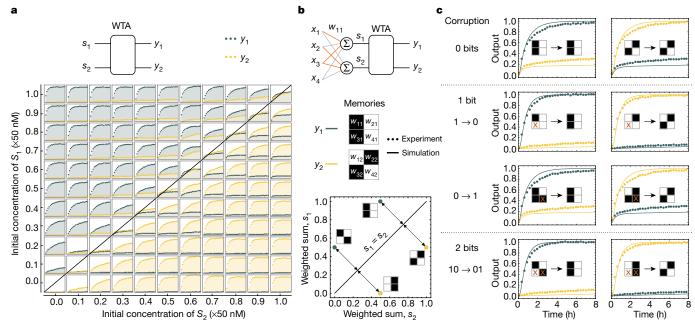


Fig. 2 | Experimental characterization of winner-take-all DNA neural networks. a, Two-species winner-take-all behaviour. The standard concentration is 50 nM (1 \times). The circuit is composed of two weightedsum strands (S_1 and S_2), an annihilator molecule ($[Anh_{1,2}] = 75$ nM $(1.5\times)$), two restoration gates ([RG₁] = [RG₂] = 50 nM (1×)), two fuel strands ($[YF_1] = [YF_2] = 100 \text{ nM } (2 \times)$) and two reporters ([Rep₁] = [Rep₂] = 100 nM (2×)). Initial concentrations of S_1 and S_2 are shown as fractions of the standard concentration. The diagonal line indicates equal concentrations of both strands. Fluorescence kinetics data are shown over the course of 2.5 h, normalized using a common minimum and maximum fluorescence level (Methods section 'Data normalization'). To clearly illustrate the difference between the two output trajectories, the background below the data points are shown in the same colour (with some transparency) as the data points. b, A 4-bit pattern-recognition circuit. In the weighted-sum layer of the circuit diagram (top left), each wire corresponds to a weight molecule, all wires from the same input require a common fuel strand and all wires to the same output require a common summation gate. Thus, a circuit that can remember any two 4-bit patterns is composed of 25 molecules (4 inputs, 14 molecules in the weighted-sum layer and 7 molecules in the winner-take-all layer).

However, a circuit that remembers two specific 4-bit patterns requires only a subset of the wires in the weighted-sum layer, each corresponding to a 1 in the memories (for example, each orange wire in the circuit diagram corresponds to a black pixel in the memories). Thus, the example circuit is composed of 20 molecules (4 fewer weight molecules and 1 fewer fuel strand). In each output-trajectory plot (right), dotted lines indicate fluorescence kinetics data and solid lines indicate simulations. The patterns to the left and right of the arrows indicate input signals and output classifications, respectively. Each orange cross indicates a bit-flip compared to the memories. The initial concentration of each input strand or weight molecule is either 0 or 50 nM; weight fuels (XF₁ and XF₂) are twice the concentration of weight molecules; the initial concentrations of the summation gates, annihilator, restoration gates, restoration fuels and reporters are 100 nM (1×), 400 nM (4×), 100 nM (1×), 200 nM (2×) and 200 nM (2×), respectively, with a standard concentration of 100 nM (details in Supplementary Table 3). In the weighted-sum space (bottom left), the two patterns with two corrupted bits are the same distance (shown as double-headed arrows) from the diagonal line as the two perfect inputs.

reporter molecule Rep_j (Extended Data Fig. 1) to separate the fluorophore- and quencher-labelled strands in the reporter, resulting in increased fluorescence. Overall, the implementation of an arbitrary winner-take-all neural network can be mapped systematically to a see-saw DNA circuit (Extended Data Fig. 2).

We started the experimental demonstration with a two-species winner-take-all function (Fig. 2a), which is similar to approximate majority¹⁷ and consensus network¹⁸ functions. If the initial concentration of one weighted-sum species (S_1 or S_2) is higher than that of the other, then we expect the corresponding output strand $(Y_1 \text{ or } Y_2)$ to be released catalytically and the fluorescent signal to reach an ideal ON state, while the other output signal remains at an ideal OFF state. The data agree with the expected overall circuit behaviour, and lead to two main observations. First, the circuit computed an ON state faster with a larger difference between the two species, as shown in the plots farther away from the diagonal line in Fig. 2a. This is because the signalrestoration reaction reaches completion faster with a larger amount of catalyst, which is the leftover amount of the winner after the annihilation reaction. Second, among experiments for which the differences between the two species are the same, the circuit maintained a cleaner OFF state with lower initial concentrations of the two species, as shown in the plots that are equidistant to the diagonal line but closer to the bottom left corner of the grid. This is because a small fraction of the weighted-sum strands will interact with a restoration-gate molecule

before encountering an annihilator molecule—the stronger the runner-up is (that is, with a higher concentration), the more it can escape the process of being completely annihilated. These observations suggest that the DNA circuit does not yield a perfect winner-take-all behaviour, but that it does compute correctly for competitors that are not too similar to each other and are not both too strong.

Next, we added a weighted-sum layer to the winner-take-all circuit to demonstrate recognition of 4-bit patterns (Fig. 2b). Using the two target patterns as weights, the perfect input patterns each triggered the desired output trajectory to turn ON, indicating that the inputs were recognized correctly. When one or two bits of the input patterns were flipped, either from a 1 to a 0 or vice versa, the circuit still yielded the desired output for all six examples that are classifiable. The other eight possible inputs are not classifiable because they result in equal weighted sums ($s_1 = s_2$). Interestingly, the circuit behaviour was better for the inputs with 2-bit corruptions than for the perfect inputs: the ON trajectories reached completion just as fast and the OFF trajectories remained lower. This result can be understood by looking at the input patterns in the weighted-sum space (Fig. 2b, bottom left): all four inputs are equidistant to the diagonal line and the corrupted patterns are closer to the bottom left corner of the space. Because catalytic reactions are used to implement weight multiplication, together with thresholding reactions, the circuit can also handle a range of input concentration that varies from the ideal high or low concentration (Extended Data Fig. 3).

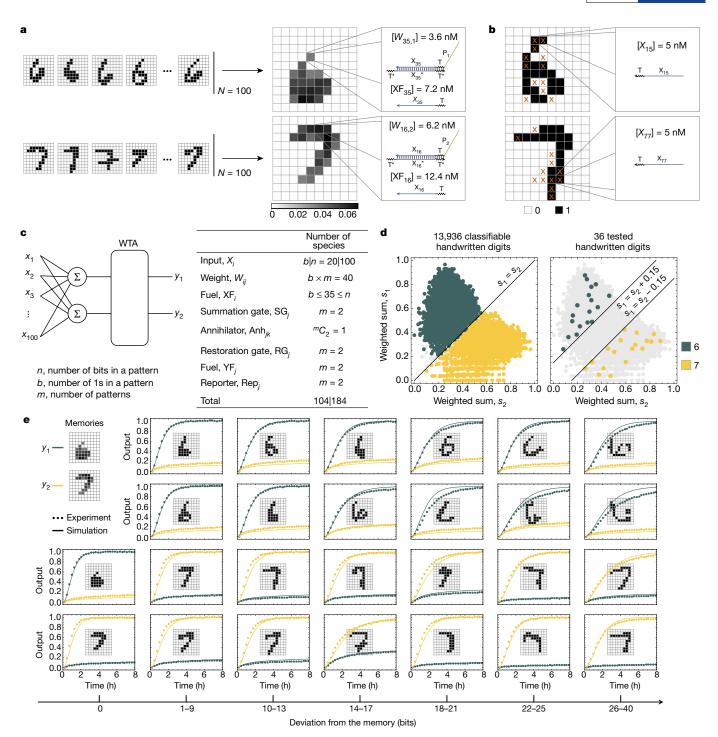


Fig. 3 | A winner-take-all DNA neural network that recognizes 100-bit patterns as one of two handwritten digits. a, Weights determined as the average of 100 '6's and '7's from the MNIST database. The value of each pixel (for example, 0.036 for the 35th pixel in '6' and 0.062 for the 16th pixel in '7') was used to determine the concentration of each weight molecule, relative to a standard concentration of 100 nM (for example, $[W_{35,1}] = 3.6$ nM and $[W_{16,2}] = 6.2$ nM). The concentrations of the fuel strands that facilitate the weight multiplication reactions were twice that of their respective weight molecules. b, Example binary inputs with each 1 and 0 corresponding to the presence and absence of an input strand, respectively. The concentration of each input strand present was $1/b \times 100 \text{ nM} = 5 \text{ nM}$, where b = 20 is the total number of 1s in each input. The orange crosses indicate bit-flips compared to the memories (that is, weight matrices) shown in a. There are 12 flipped bits in each example. Because the total number of 1s in each input pattern is the same as the total number of non-zero weights in the memories, it is always the case that half of the flipped bits are associated with non-zero weights. c, Circuit diagram

and the number of distinct species in the circuit. For the total number of species, the two values correspond to the number of species for a specific number b of inputs (left) and for all n possible inputs (right). **d**, The 13,936 classifiable digits (left; large green and yellow points) correspond to 98% of all '6's and '7's in the MNIST database. Test input patterns were chosen (right; large green and yellow points; 36 in total) on the basis of their locations in the weighted-sum space. The lines labelled $s_1 = s_2 \pm 0.15$ indicate a 15% margin to the diagonal line, within which we expect the pattern recognition to be experimentally difficult. The light grey points correspond to non-classifiable (left) or non-tested (right) digits. e, Recognizing handwritten digits with up to 30 flipped bits compared to the 'remembered' digits. Dotted lines indicate fluorescence kinetics data and solid lines indicate simulations. The input pattern is shown in each plot. Note that 40 is the maximum number of flipped bits because all patterns have exactly 20 1s. Weights and inputs are listed in Supplementary Table 2. The initial concentrations of all species are listed in Extended Data Fig. 10 (details in Supplementary Table 3).

To understand the theoretical limits of the scalability and power of winner-take-all DNA neural networks, in the context of simply using the target patterns as weights, we now address the following three questions. The first is the number of distinct target patterns that can be remembered simultaneously. Any set of patterns that consists of the same number of 1s can be remembered (Methods, Theorem 1). For example, the largest set of 9-bit patterns that can be remembered, each consisting of five 1s, consists of ${}^{9}C_{5} = 126$ patterns. Moreover, any set of patterns can be remembered if it does not contain a pattern in which all 1s are a subset of 1s in another pattern (Methods, Theorem 2). The second question concerns which corrupted patterns can be recognized. All patterns with fewer than b - o corrupted bits can be recognized, where *b* is the total number of 1s and *o* is the maximum number of overlapped 1s in all target patterns (Methods, Theorem 3). For example, all patterns with fewer than three corrupted bits can be recognized for the 9-bit target patterns 'L' and 'T' shown in Fig. 1b, because b = 5 and o = 2. Moreover, some patterns with more than b - o corrupted bits can still be recognized; for example, in all possible 9-bit patterns, there are 128, 102 and 30 patterns with three, four and five corrupted bits, respectively, that can be recognized as 'L' or 'T'. We chose 28 example 9-bit patterns with an increasing number of corrupted bits from one to five, and demonstrated that the DNA neural network correctly classified all examples (Extended Data Fig. 4). The final question asks how the size of the DNA circuit scales with an increasing number of more complex patterns. In general, constructing a network that can remember m distinct *n*-bit patterns requires *n* input strands, $n \times m$ weight molecules and n fuel strands for weight multiplication, m summation gates, ${}^{m}C_{2}$ annihilators, m gates and m fuel strands for signal restoration, and m reporters, totalling $n \times m + 2n + 4m + {}^{m}C_{2}$ molecules. However, for a specific set of target patterns, only a subset of the weight molecules are required, each corresponding to a 1 in the patterns.

To demonstrate the scalability and power of winner-take-all DNA neural networks experimentally, we chose a task that is visually interesting: recognizing handwritten digits. Some aspects of this task are computationally non-trivial, such as distinguishing a sloppy '4' from a sloppy '9'. The patterns of digits were taken from the Modified National Institute of Standards and Technology (MNIST) database¹⁹, which is commonly used to test machine learning algorithms²⁰. We converted the original patterns to binary patterns with 20 1s on a 10×10 grid, averaged 100 example '6' and '7' patterns, and selected and normalized the top 20 pixels as weights (Fig. 3a, Methods section 'Neural network training and testing'). The value of each analogue weight was then implemented with the concentration of a weight molecule. The test inputs remained binary patterns, in which each 1 or 0 corresponded to the presence or absence of an input strand, respectively (Fig. 3b). The theoretical limits of the winner-take-all neural networks with analogue weights are similar to those with binary weights (Methods, Theorems 4 and 5). In total, 104 distinct molecules were used for testing any specific input pattern out of 184 distinct molecules for all possible inputs (Fig. 3c).

In the MNIST database, there are more than 14,000 example handwritten '6' and '7' digits. On the basis of the understanding that we have established from the experimental characterization of smaller winnertake-all circuits, we looked at all example patterns in the weighted-sum space (Fig. 3d, Extended Data Fig. 5a): 2% of the patterns are on the wrong side of the diagonal line, which means that it is impossible for the DNA circuit to recognize them correctly; 8% of the patterns are fairly close to the diagonal line (within a 15% margin), which we expect to be experimentally difficult; however, the remaining 90% of the patterns are far enough from the diagonal line that we expect correct recognition. Therefore, we chose 36 representative example patterns from the last category, ensuring both uniform distribution in the weighted-sum space and the full range of bit deviation from the memories (Methods section 'Neural network training and testing'). As shown in the experimental data (Fig. 3e, Extended Data Fig. 5d), the perfect patterns (the weights converted to binary) each yield the desired circuit output. More importantly, patterns that increasingly deviate from the memories

were also recognized, with up to 30 flipped bits. Similar to observations in the smaller DNA neural networks, some of the patterns that are visually more challenging to recognize are not necessarily more difficult for the DNA circuit—a desirable property of the winner-take-all computation.

We have shown that the winner-take-all DNA neural networks scale well to more complex patterns. Next, we explore whether they could also be used to remember an increasing number of distinct patterns simultaneously. The pairwise-annihilation approach alone is not well suited for scaling up the number of patterns because the number of annihilators grows quadratically with the number of patterns. We show that the three-species winner-take-all function was still robust enough (Extended Data Fig. 6a) to allow the construction of a DNA neural network that remembers three 100-bit patterns. However, the competition became harder with more competitors: the reaction rates for multiple annihilation pathways could be matched approximately but not perfectly (Methods section 'Sequence design', Extended Data Fig. 6b, c), and it took much longer for the annihilation reactions to yield a winner and for the signal level of the winner to be fully restored (Extended Data Fig. 7). Using the same method, it would be difficult to construct networks that remember more patterns. We therefore propose an alternative approach that first divides the target patterns into groups and then uses multiple distinct group identities to classify the patterns (Fig. 4a). The nine digits '1'-'9' can be divided into three groups in two ways (shown as three rows and three columns in Fig. 4b), such that a pair of outputs corresponds uniquely to each digit (Fig. 4d). For example, a '4' is recognized if and only if $y_1 = 1$ and $z_1 = 1$ (where y_1 is the output identifying the first row and z_1 is the output identifying the first column). With this grouping approach, nine distinct patterns can be recognized using only $\sqrt{9}\bar{C}_2 \times 2 = 6$ annihilators, which would otherwise require ${}^{9}C_{2} = 36$ annihilators. In total, 225 distinct molecules were used for testing any specific input pattern out of 305 distinct molecules for all possible inputs (Fig. 4c).

We determined the weights for each group using a simple 'average then subtract' method (Fig. 4b): take the average of 100 examples per in-group digit, subtract the average of 100 examples per out-of-group digit, then select and normalize the top 20 pixels (Methods section 'Neural network training and testing'). The trade-off of the grouping approach is that fewer example patterns can be recognized. With the best grouping, 47% of the patterns can potentially be recognized, of which 48% are experimentally feasible (with a 15% margin to the diagonal line in the normalized weighted-sum space). In general, with the same circuit complexity, this alternative approach enables a larger set of distinct target patterns to be classified, but with less accuracy. Nonetheless, as shown in the experimental data, the circuit yields the desired pair of outputs for 99 representative example patterns (Fig. 4d, e).

To facilitate the design of winner-take-all DNA neural networks, we developed an online software tool. The WTA Compiler²¹ (Extended Data Fig. 8) converts a user-defined set of memories and test patterns into program code that describes a DNA neural network, which can then be used to simulate the kinetics of the network. It also provides sequences of the DNA strands that are required to construct the DNA neural network experimentally.

It is interesting to compare the performance of winner-take-all neural networks with logic circuits. For example, it is possible to distinguish whether a 9-bit pattern is more similar to 'L' or 'T' using a circuit consisting of 8 logic gates, for all input patterns that we have tested experimentally. However, a more complex circuit consisting of 21 logic gates is required to correctly compute the output for all classifiable patterns (Extended Data Fig. 9a). Similarly, the 100-bit handwritten digits can be recognized by circuits with up to 23 logic gates, if only the example patterns that we have tested experimentally are considered. But these logic circuits perform poorly when tested against the entire MNIST database (Extended Data Fig. 9b). To match the theoretical limit of winner-take-all neural networks, measured by the percentage of classifiable patterns, much more complex logic circuits are needed. Importantly,

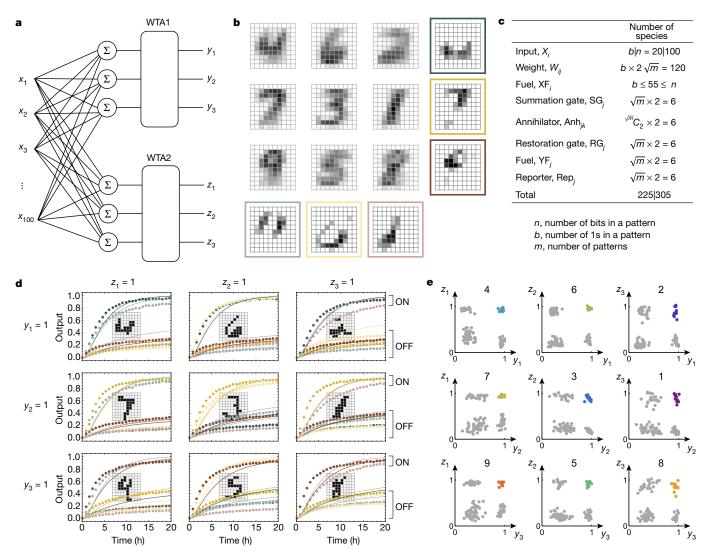


Fig. 4 | **A winner-take-all DNA neural network that recognizes 100-bit patterns as one of nine handwritten digits. a**, Circuit diagram for recognizing nine distinct patterns using a grouping approach. WTA1 and WTA2 are two separate winner-take-all functions that each yields a distinct set of outputs $(y_j$ and $z_j)$. **b**, Weights determined using an 'average then subtract' method. The average of 100 example digits from the MNIST database are shown grouped by rows (corresponding to outputs y_j) and columns (corresponding to outputs z_j). The weight matrix for each group (boxed patterns; colours correspond to the respective output trajectories in **d**) is the average of all in-group digits less the average of all out-of-group digits. Using this weight matrix, the fraction of experimentally feasible test patterns from all examples in the MNIST database was calculated for all possible ways of grouping the nine digits and the best grouping was chosen

and shown here. **c**, Number of distinct species in the circuit in **a**. For the total number of species, the two values correspond to the number of species for a specific number b of inputs (left) and for all n possible inputs (right). **d**, Fluorescence kinetics data (dotted lines) and simulations (solid lines) of the circuit behaviour with nine representative input patterns (shown in the plots). **e**, Fluorescence level of each pair of outputs at 24 h or longer after the inputs were added, collected from 99 experiments with 11 example patterns per digit. Each coloured point corresponds to an example pattern from the labelled class of digit; each grey point corresponds to an out-of-class example pattern. Weights and inputs are listed in Supplementary Table 2. The initial concentrations of all species are listed in Extended Data Fig. 10 (details in Supplementary Table 3).

varying the concentrations of the weight molecules in the winner-takeall neural networks would enable the same set of DNA molecules to be used for different pattern-classification tasks. By contrast, without reconfigurable circuit architectures, a different set of DNA molecules would be required for a logic circuit that performs a different task.

The power of winner-take-all DNA neural networks could be explored further in several directions. Instead of the pairwise-annihilation approach, a winner could be selected by utilizing competing resources^{5,6}, which could potentially lead to more scalable and accurate pattern recognition. It could also provide the possibility of selecting several winners instead of just one, which in theory is computationally more powerful⁴. Extending the circuit construction from single-layer to multi-layer winner-take-all computation, or simply allowing the outputs of winner-take-all circuits to be connected to downstream logic circuits, could enable more sophisticated pattern

recognition (such as involving translated and rotated patterns)²². Using a variable-gain amplifier^{23,24}, winner-take-all DNA circuits could be adapted to process analogue inputs, which would enable a wider range of signal-classification tasks, including applications in detecting complex disease profiles that consist of mRNA and microRNA signals. With aptamers^{25,26}, more diverse biomolecules could be detected.

The fact that we were able to use target patterns as weights in winner-take-all DNA neural networks opens up immediate possibilities for embedding learning within autonomous molecular systems. With one additional circuit component that actives weight molecules during a supervised training process, the DNA circuits would be capable of activating a specific set of wires in the weight-multiplication layer when exposed to a specific set of patterns. As widely discussed in experimental²⁷ and theoretical^{28–30} studies, learning—the most desirable property



of biochemical circuits—would allow artificial molecular machines to adapt their functions on the basis of environmental signals during autonomous operations.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0289-6.

Received: 30 October 2017; Accepted: 18 April 2018; Published online 4 July 2018.

- Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. 5, 1024–1037 (2004).
- Mori, K., Nagao, H. & Yoshihara, Y. The olfactory bulb: coding and processing of odor molecule information. Science 286, 711–715 (1999).
- Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* 475, 368–372 (2011).
- Maass, W. On the computational power of winner-take-all. Neural Comput. 12, 2519–2535 (2000).
- Kim, J., Hopfield, J. & Winfree, E. Neural network computation by in vitro transcriptional circuits. Adv. Neural Inf. Process. Syst. 17, 681–688 (2005)
- Genot, A. J., Fujii, T. & Rondelez, Y. Scaling down DNA circuits with competitive neural networks. J. R. Soc. Interface 10, 20130212 (2013).
- Muroga, S. Threshold Logic and its Applications (Wiley Interscience, New York, 1971).
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl Acad. Sci. USA 79, 2554–2558 (1982).
- Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* 406, 605–608 (2000).
- Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using stranddisplacement reactions. *Nat. Chem.* 3, 103–113 (2011).
- Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. Science 332, 1196–1201 (2011).
- Thubagere, A. J. et al. Compiler-aided systematic construction of large-scale DNA strand displacement circuits using unpurified components. *Nat. Commun.* 8, 14373 (2017).
- Zhang, D. Y. Cooperative hybridization of oligonucleotides. J. Am. Chem. Soc. 133, 1077–1086 (2011).
- Redgrave, P., Prescott, T. J. & Gurney, K. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89, 1009–1023 (1999).
- Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. J. Am. Chem. Soc. 131, 17303–17314 (2009).
- Yurke, B. & Mills, A. P. Using DNA to power nanostructures. Genet. Program. Evol. Mach. 4, 111–122 (2003).
- Cardelli, L. & Csikász-Nagy, A. The cell cycle switch computes approximate majority. Sci. Rep. 2, 656 (2012).
- Chen, Y.-J. et al. Programmable chemical controllers made from DNA. Nat. Nanotechnol. 8, 755–762 (2013).

- LeCun, Y., Cortes, C. & Burges, C. J. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/index.html.
- Deng, L. The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Process. Mag.* 29, 141–142 (2012).
- Cherry, K. M. WTA Compiler. http://www.qianlab.caltech.edu/WTAcompiler/ (2017).
- Rojas, R. Neural Networks: A Systematic Introduction (Springer, Berlin, 2013).
 Zhang, D. Y. & Seelig, G. DNA-based fixed gain amplifiers and linear classifier circuits. In DNA 2010: DNA Computing and Molecular Programming (eds Sakakibara, Y. & Mi, Y.) 176–186 (Springer, 2011).
- Chen, S. X. & Seelig, G. A DNA neural network constructed from molecular variable gain amplifiers. In DNA 2017: DNA Computing and Molecular Programming (eds Brijder, R. & Qian, L.) 110–121 (Springer, Cham, 2017).
- Cho, E. J., Lee, J.-W. & Ellington, A. D. Applications of aptamers as sensors. Annu. Rev. Anal. Chem. 2, 241–264 (2009).
- Li, B., Ellington, A. D. & Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Res.* 39, e110 (2011).
- 27. Pei, R., Matamoros, E., Liu, M., Stefanovic, D. & Stojanovic, M. N. Training a molecular automaton to play a game. *Nat. Nanotechnol.* **5**, 773–777 (2010).
- Fernando, C. T. et al. Molecular circuits for associative learning in single-celled organisms. J. R. Soc. Interface 6, 463–469 (2009).
- Aubert, N. et al. Evolving cheating DNA networks: a case study with the rock–paper–scissors game. In ECAL 2013: Advances in Artificial Life (eds Liò, P. et al.) 1143–1150 (MIT Press, Cambridge, 2013).
- Lakin, M. R., Minnich, A., Lane, T. & Stefanovic, D. Design of a biochemical circuit motif for learning linear functions. J. R. Soc. Interface 11, 20140902 (2014).

Acknowledgements We thank R. M. Murray for sharing an acoustic liquid-handling robot. We thank C. Thachuk and E. Winfree for discussions and suggestions. K.M.C. was supported by a NSF Graduate Research Fellowship. L.Q. was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (1010684), a Faculty Early Career Development Award from NSF (1351081), and the Shurl and Kay Curci Foundation.

Reviewer information *Nature* thanks R. Schulman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.M.C. developed the model, designed and performed the experiments, and analysed the data; K.M.C. and L.Q. wrote the manuscript; L.Q. initiated and guided the project.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0289-6.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.Q. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Sequence design. All DNA strands used in the winner-take-all neural networks were composed of long branch-migration domains and short toehold domains. Owing to the modularity of the previously developed seesaw DNA motif^{3,11} and the extended new circuit component—the annihilator—the sequence design was performed at the domain level. A pool of domain sequences was generated according to a set of design heuristics that have previously been experimentally validated 12. All domains used a three-letter code (A, T and C) to reduce secondary structure and undesired strand interactions. No domain sequences include runs of more than four consecutive As or Ts or more than three consecutive Cs, which reduces synthesis errors. All domain sequences had between 30% and 70% C-content so all double-stranded complexes would have similar melting temperatures. Finally, no pairs of domain sequences share a matching sequence longer than 35% of the domain length, and all pairs have at least 30% different nucleotides. This ensures that a strand with a mismatched branch-migration domain will not complete strand displacement initiated from either the 3' or the 5' end. In addition to a 15-nucleotide sequence pool used in previous work^{3,11,12}, a 20-nucleotide sequence pool was generated and used in the weight multiplication layers because of the large number of molecules used here. The two sequence pools were checked to ensure that the same pairwise criteria were met. All domains included the clamp design introduced previously¹¹, to reduce leak reactions between initial gate species.

All molecular complexes shared a 5-nucleotide universal toehold domain^{3,11,12}. The annihilator complexes had 7-nucleotide toeholds composed of the 5-nucleotide universal toehold and a 2-nucleotide extension that matched the 2 nucleotides adjacent to the toehold on the upstream seesaw gate. This increased the binding energy and thus the effective strand-displacement reaction rate between the annihilator complexes and the weighted-sum strands, compared to that between the signal-restoration gates and the weighted-sum strands.

To ensure 'fair competition' between the weighted-sum species (that is, same rates for all pairwise-annihilation reactions), all annihilators within a set of winner-take-all computations had identical toehold extensions, and the weightedsum strands had the same single-nucleotide dangle to keep the binding energies consistent within a winner-take-all computation. Here, we used up to two sets of three annihilators. The extensions and dangle sequences were chosen by estimating the binding energies using NUPACK³¹, and the sequences for the second set of annihilators were chosen with similar energies to those of the first set that worked well in the three-species winner-take-all experiments (Extended Data Fig. 6a). In addition, the rate of an annihilation reaction could depend on the sequence of the branch-migration domains. We measured the rates of 15 catalytic gates, and selected two groups of three gates with the closest rates (Extended Data Fig. 6b, c). By using these gates for signal restoration, the branch-migration domains in the annihilators were determined simultaneously, because the signal-restoration gates and annihilators share the same branch-migration domains (Extended Data Fig. 1). All DNA sequences are listed in Supplementary Table 1.

Neural-network training and testing. The winner-take-all DNA neural network was tested on patterns derived from the MNIST handwritten-digit database 19 . The training and testing sets were downloaded and merged into a single database, and all example patterns of digits '1'–'9' were retained, totalling 63,097 images. The original MNIST dataset consists of weight-centred grey-scale images on a 28×28 grid. Here, we used binary patterns on a 10×10 grid. First, the images were rescaled to a 12×12 grid using Gaussian resampling. The largest 20 bits in each image were set to 1 and the remaining bits were set to 0. Finally, the digits were re-centred on a 10×10 grid on the basis of their bounding boxes.

We made a conscious effort to train the neural networks using a simple algorithm. In the neural networks that remember two or three handwritten digits, for each digit, the weight matrices were the average of the first 100 example patterns in the database, restricted to the 20 most common bits (that is, the ones with the largest averaged values), and normalized to sum to 1. For the nine-digit network, all digits were divided into three groups in two ways. For each group, the weight matrix was the average of the first 100 examples of the three in-group digits less the average of the first 100 examples of the six out-of-group digits. The 20 most common bits were retained, and all weight matrices were normalized to sum to 1.15, to shift the test patterns into a more ideal area in the weighted-sum space. The fraction of experimentally feasible test patterns (with a 15% margin to the diagonal line in the weighted-sum space for all pairs of species) was calculated for all ways of grouping the nine digits, and the best grouping was chosen. The classification performance of the network using weights determined by non-negative least squares was only slightly better than the performance using weights from the simple 'average then subtract' method (54% versus 47%).

Experimentally tested input patterns were chosen to represent the whole weightedsum space as well as the full range of bit deviation from the memories of the networks. To choose a set of test patterns for a digit, all correctly classified examples of that digit with at least a 15% margin in the weighted-sum space were divided into six corruption classes. The weighted sums for the digits in each class were then clustered using the k-medoids algorithm, and an example test pattern was chosen randomly from each cluster according to a uniform distribution. This ensured that the test patterns represented the whole weighted-sum space and not just the most common digits.

Weights and inputs used in all experiments are listed in Supplementary Table 2. By exporting each sheet of the Excel file to a .csv file and uploading it to the WTA Compiler²¹, the weights and inputs can be visually displayed, the inputs analysed in their weighted-sum space, the kinetics behaviour of the winner-take-all DNA neural network simulated and DNA sequences generated.

DNA oligonucleotide synthesis. All DNA strands were purchased from Integrated DNA Technologies (IDT). The reporter strands with fluorophores and quenchers were purified (HPLC) and the other strands were unpurified (standard desalting). All strands were shipped lyophilized then resuspended at $100~\mu M$ in Tris-EDTA (TE) buffer, pH 8.0, and stored at $4~\rm ^{\circ}C$.

Annealing protocol and buffer condition. Annihilator and gate complexes were prepared for annealing at 45 μ M with top and bottom strands in a 1:1 ratio. Reporters were prepared at 20 μM with top quencher strands in 20% excess of bottom strands. The buffer for all experiments and annealed complexes was TE with 12.5 mM Mg²⁺. Complexes were annealed in a thermal cycler (Eppendorf) by heating to 90 °C for 5 min and then cooling to 20 °C at a rate of 0.1 °C per 6 s. Purification. Annealed annihilator and gate complexes were purified using 12% polyacrylamide gel electrophoresis (PAGE). Double-stranded complex bands were cut from the gel, chopped into pieces and incubated for 24 h at room temperature in TE buffer with 12.5 mM Mg²⁺ to allow DNA to diffuse into the buffer. The solution with purified complexes was recovered and concentrations were determined with NanoDrop (Thermo Fisher). Weight matrices for the DNA neural networks that remember handwritten digits had 20 gate complexes for each neuron. These gates (weight molecules) were annealed individually and then mixed together in the appropriate ratio, on the basis of the values of the weights. This mixture was then purified via PAGE, recovered and the concentration determined by NanoDrop

using the weighted-average extinction coefficient. Fluorescence spectroscopy. Fluorescence kinetics data were collected every 2, 3 or 4 min, depending on the overall length of the experiment, using a microplate reader (Synergy H1, Biotek). Excitation (emission) wavelengths were 496 nm (525 nm) for dye ATTO488, 555 nm (582 nm) for dye ATTO550 and 598 nm (629 nm) for dye ATTO590. Experiments were performed in 96-well plates (Corning) with 160-µl reaction mixture per well for the nine-digit experiments and 200-µl reaction mixture per well for all other experiments. Experiments were performed at a standard concentration of 100 nM for all 4-bit and 100-bit pattern recognition and at a standard concentration of 50 nM for all other experiments. Initial concentrations of all species are listed in Extended Data Fig. 10. Detailed protocols for all experiments are listed in Supplementary Table 3.

In the nine-digit experiments, six distinct output trajectories were read using three distinct fluorophores. Every experiment was run twice, each having half of the outputs connected to fluorophore-labelled reporters and the other half to non-fluorophore-labelled reporters. Combining the output trajectories from each pair of experiments into a single plot allows the observation of all six outputs simultaneously.

Data normalization. All data were normalized from raw fluorescence level to standard concentration, which is the maximum concentration of an output strand Y_j released from gate RG_j and interacted with a double-stranded reporter molecule Rep_j . The fluorescence level that corresponds to standard concentration $(1\times)$ was obtained from the average of the final five measurements from the highest signal produced from gate RG_j on a plate. Negligible concentration $(0\times)$ corresponds to the background fluorescence of the reaction mixture before any reporter molecules have been triggered, which was obtained from the first measurement of the lowest signal produced from gate RG_j on a plate. All experiments on a single plate were normalized together, allowing direct comparison between the output of a network for different input patterns. In the two-species winner-take-all experiments shown in Extended Data Fig. 3, the first six columns of data were measured on one plate and the last five columns measured on another. In the 9-bit pattern-recognition experiments shown in Extended Data Fig. 4, the input patterns with 0–2 corrupted bits were measured on one plate and those with 3–5 corrupted bits were measured on another.

Model and simulations. Mass-action simulation were performed using the same set of reactions and rate constants developed in the seesaw model¹¹, with four additional reactions to model pairwise annihilation:

$$S_{j} + \operatorname{Anh}_{jk} \stackrel{k_{f}}{\rightleftharpoons} S_{j} \cdot \operatorname{Anh}_{jk}$$

$$S_{k} + \operatorname{Anh}_{jk} \stackrel{k_{f}}{\rightleftharpoons} S_{k} \cdot \operatorname{Anh}_{jk}$$

$$S_{j} \cdot \operatorname{Anh}_{jk} + S_{k} \stackrel{k_{f}}{\rightleftharpoons} \varnothing$$

$$S_{k} \cdot \operatorname{Anh}_{jk} + S_{j} \stackrel{k_{f}}{\rightleftharpoons} \varnothing$$

Here, $k_{\rm f}=2\times 10^6\,{\rm M}^{-1}{\rm s}^{-1}$, which is the same as the forward rate constant of the thresholding reaction in the seesaw model¹¹. The reverse rate constant $k_{\rm r}=0.4\,{\rm s}^{-1}$ was determined using the experimental data shown in Extended Data Fig. 3a. This rate constant is of the same order as found in a previous study of cooperative hybridization¹³. Similar to the spurious reactions in the original seesaw model, temporary toehold binding between any single-stranded species and any annihilator (or intermediate annihilator species listed above) are also included here. Code availability. Simulation code is available at the WTA Compiler website²¹. Theoretical limits of the power of winner-take-all neural networks. The winner-

Definition 1. Let $X = \{x^1, x^2, ..., x^m\}$ be a set of m patterns, each with n bits. Let an example pattern from X be $\mathbf{x}^{\alpha} = (x_1^{\alpha}, x_2^{\alpha}, \cdots, x_n^{\alpha})$, with $x_i^{\alpha} \in \{0, 1\}$. We say that a winner-take-all neural network with weights W remembers X if $y_{\alpha} = 1$ for all $1 \le \alpha \le m$ (and $y_i = 0$ for all $j \ne \alpha$) when $\mathbf{x} = \mathbf{x}^{\alpha}$.

Theorem 1. If X is a set of m distinct n-bit patterns, each containing exactly b 1s, then the winner-take-all neural network with $W = (\mathbf{w}_1^{\top}, \mathbf{w}_2^{\top}, ..., \mathbf{w}_m^{\top})$ and $\mathbf{w}_j = (\mathbf{w}_{1j}, \mathbf{w}_{2j}, ..., \mathbf{w}_{nj}) = \mathbf{x}^j$ (that is, $\mathbf{w}_{ij} = \mathbf{x}_i^j$) remembers X.

Proof. Consider this network on input $\mathbf{x} = \mathbf{x}^{\alpha}$. First, for $j = \alpha$, we calculate $s_{\alpha} = \mathbf{x}^{\alpha} \cdot \mathbf{x}^{\alpha} = b$. Second, for $j \neq \alpha$, $\mathbf{x}^{j} \neq \mathbf{x}^{\alpha}$. Because the number of 1s in both of these patterns is b, the number of indices at which the bits are both 1 is strictly less than b. Therefore, $s_{j} = \mathbf{x}^{j} \cdot \mathbf{x}^{\alpha} < b$. Putting the first and second calculations together, we conclude that $s_{\alpha} > s_{j}$ and thus $y_{\alpha} = 1$ and $y_{j} = 0$ for all $j \neq \alpha$.

The next theorem is a generalization of Theorem 1.

take-all function shown in Fig. 1a is defined to have:

Theorem 2. If X is a set of m distinct n-bit patterns, and the 1s in any example pattern \mathbf{x}^{α} is not a subset of the 1s in another pattern \mathbf{x}^{β} (that is, no two example patterns satisfy $\mathbf{x}^{\alpha} \cdot \mathbf{x}^{\beta} = \mathbf{x}^{\alpha} \cdot \mathbf{x}^{\alpha}$), then the winner-take-all neural network with $W = (\mathbf{w}_{1}^{\top}, \mathbf{w}_{2}^{\top}, ..., \mathbf{w}_{m}^{\top})$ and $\mathbf{w}_{j} = \mathbf{x}^{j}$ remembers X.

Proof. Consider this network on input $\mathbf{x} = \mathbf{x}^{\alpha}$. First, $s_{\alpha} = \mathbf{x}^{\alpha} \cdot \mathbf{x}^{\alpha}$ and is equal to the total number of 1s in \mathbf{x}^{α} . Second, for $j \neq \alpha$, $s_{j} = \mathbf{x}^{j} \cdot \mathbf{x}^{\alpha} \neq \mathbf{x}^{\alpha} \cdot \mathbf{x}^{\alpha}$. Third, for all j, $s_{j} = \mathbf{x}^{j} \cdot \mathbf{x}^{\alpha} \leq \mathbf{x}^{\alpha} \cdot \mathbf{x}^{\alpha} = s_{\alpha}$. Putting these three constraints together, we conclude that $s_{\alpha} > s_{j}$ and thus $y_{\alpha} = 1$ and $y_{j} = 0$ for all $j \neq \alpha$.

Definition 2. In a winner-take-all neural network with $W = (w_1^\top, w_2^\top, ..., w_m^\top)$ and $w_j = x^j$, we that say each x^j is a *memory*. We say that the network *recognizes* input x as memory x^α if $y_\alpha = 1$ (and $y_j = 0$ for all $j \neq \alpha$). We say that a pattern x has c corrupted bits compared to a memory x^α (or has c-bit deviation from x^α) if the number of indices at which the bits are different (that is, one bit is 0 and the other is 1 or vice versa) in x and x^α is exactly c. We say that two memories x^α and x^β have o overlapped bits if the number of indices at which the bits are both 1 in these memories is exactly o.

Theorem 3. If x is a pattern with c < b - o corrupted bits compared to a memory x^{α} , where b is the total number of 1s in x^{α} and o is the maximum number of overlapped bits in x^{α} and x^{j} for all $j \neq \alpha$, then the winner-take-all neural network recognizes x as x^{α} .

Proof. Let c_0 be the number of flipped 0s (that is, where 1 in x and 0 in x^{α} appear at the same index) and c_1 be the number of flipped 1s (that is, where 0 in x and 1 in x^{α} appear at the same index). First, $s_{\alpha} = x^{\alpha} \cdot x = b - c_1$. Second, for $j \neq \alpha$,

 $s_j = \mathbf{x}^j \cdot \mathbf{x} \le o + c_0$ (s_j reaches its maximum when all corrupted 1s are 0s and all corrupted 0s are 1s are at the same indices in \mathbf{x}^j). Third, because $c = c_0 + c_1$ and c < b - o, $o + c_0 = o + c - c_1 < o + b - o - c_1 = b - c_1$. Putting the three constraints together, we conclude that $s_\alpha > s_j$ and thus $y_\alpha = 1$ and $y_i = 0$ for all $j \ne \alpha$.

Next, we consider a much larger set of *n*-bit patterns, $X = \{x^1, x^2, ..., x^M\}$ with $M \gg m$.

Definition 3. Let each example pattern $\mathbf{x}^{\mu} = (x_1^{\mu}, x_2^{\mu}, \cdots, x_n^{\mu})$ be associated with a desired output $\mathbf{y}^{\mu} = (y_1^{\mu}, y_2^{\mu}, \cdots, y_n^{\mu})$, with $y_j^{\mu} \in \{0, 1\}$ and $\sum_{j=1}^{m} y_j^{\mu} = 1$ (that is, only one specific $y_{\alpha}^{\mu} = 1$ and $y_j^{\mu} = 0$ for all $j \neq \alpha$). If $y_{\alpha}^{\mu} = 1$, then we say that \mathbf{x}^{μ} is a pattern in *class* α .

Let $\tilde{\boldsymbol{x}}^{\alpha} = (\tilde{\boldsymbol{x}}_{1}^{\alpha}, \tilde{\boldsymbol{x}}_{2}^{\alpha}, \cdots, \tilde{\boldsymbol{x}}_{n}^{\alpha}) = \left(\sum_{\mu} \boldsymbol{x}_{1}^{\mu}, \sum_{\mu} \boldsymbol{x}_{2}^{\mu}, \cdots, \sum_{\mu} \boldsymbol{x}_{n}^{\mu}\right)$ for all μ with $\boldsymbol{y}_{n}^{\mu} = 1$ (that is, the sum of all patterns in class α). Let $t_{\alpha} = \sum_{i} \tilde{\boldsymbol{x}}_{i}^{\alpha}$ for the b largest components of $\tilde{\boldsymbol{x}}^{\alpha}$. Let $\tilde{\boldsymbol{x}}^{\alpha} = (\tilde{\boldsymbol{x}}_{1}^{\alpha}, \tilde{\boldsymbol{x}}_{2}^{\alpha}, \cdots, \tilde{\boldsymbol{x}}_{n}^{\alpha})$, with $\tilde{\boldsymbol{x}}_{i}^{\alpha} = \tilde{\boldsymbol{x}}_{i}^{\alpha}/t_{\alpha}$ if $\tilde{\boldsymbol{x}}_{i}^{\alpha}$ is one of the b largest values and $\tilde{\boldsymbol{x}}_{i}^{\alpha} = 0$ otherwise (that is, the averaged pattern for class α , restricted to the b most common bits and normalized to sum to 1). Let $\hat{\boldsymbol{x}}^{\alpha} = (\hat{\boldsymbol{x}}_{1}^{\alpha}, \hat{\boldsymbol{x}}_{2}^{\alpha}, \cdots, \hat{\boldsymbol{x}}_{n}^{\alpha})$, with $\hat{\boldsymbol{x}}_{i}^{\alpha} = 1$ if $\tilde{\boldsymbol{x}}_{i}^{\alpha} > 0$ and $\hat{\boldsymbol{x}}_{i}^{\alpha} = 0$ if $\tilde{\boldsymbol{x}}_{i}^{\alpha} = 0$. Let $\hat{\boldsymbol{X}} = \{\hat{\boldsymbol{x}}^{1}, \hat{\boldsymbol{x}}^{2}, \cdots, \hat{\boldsymbol{x}}^{m}\}$ be the set of averaged patterns converted to binary.

The next two theorems are similar to Theorems 1 and 3, but generalized to using averaged training patterns as analogue weights rather than using a single training pattern (that is, target pattern) as binary weights.

Theorem 4. If X is a set of M distinct n-bit patterns, \hat{x}^j contains exactly b 1s for all j and $\hat{x}^j \neq \hat{x}^k$ for all $j \neq k$, then the winner-take-all neural network with $W = (\mathbf{w}_1^\top, \mathbf{w}_2^\top, ..., \mathbf{w}_m^\top)$ and $\mathbf{w}_j = \overline{x}^j$ remembers \hat{X} .

Proof. Consider this network on input $\mathbf{x} = \hat{\mathbf{x}}^{\alpha}$. First, we calculate $s_{\alpha} = \overline{\mathbf{x}}^{\alpha} \cdot \hat{\mathbf{x}}^{\alpha} = \sum_{i=1}^{n} \overline{x}_{i}^{\alpha} = 1$. Second, for $j \neq \alpha$, $\hat{\mathbf{x}}^{j} \neq \hat{\mathbf{x}}^{\alpha}$. Because the number of 1s in both of these patterns is b, there exist at least one index i at which $\hat{x}_{i}^{j} = 1$ (and $\overline{x}_{i}^{j} > 0$) and $\hat{x}_{i}^{\alpha} = 0$; thus $s_{j} = \overline{\mathbf{x}}^{j} \cdot \hat{\mathbf{x}}^{\alpha} < \sum_{i=1}^{n} \overline{x}_{i}^{j} = 1$. Putting the two constraints together, we conclude that $s_{\alpha} > s_{j}$ and thus $y_{\alpha} = 1$ and $y_{\alpha} = 0$ for all $j \neq \alpha$.

Definition 4. In a winner-take-all neural network with $W = (w_1^T, w_2^T, ..., w_m^T)$ and $w_j = \bar{x}^j$, we say that each \bar{x}^j is a *memory* and each $x = \hat{x}^j$ is a *perfect* input. We say that a binary pattern x has c-bit deviation from a memory \bar{x}^α if the number of indices at which the bits are different in x and \hat{x}^α is exactly c. We say that two memories \bar{x}^α and \bar{x}^β have overlap $o = \max\{\bar{x}^\alpha \cdot \hat{x}^\beta, \bar{x}^\beta \cdot \hat{x}^\alpha\}$. We say a bit i is no more than average in \bar{x}^α if $\bar{x}^\alpha_i \leq 1/b$, where b is the total number of 1s in \hat{x}^α .

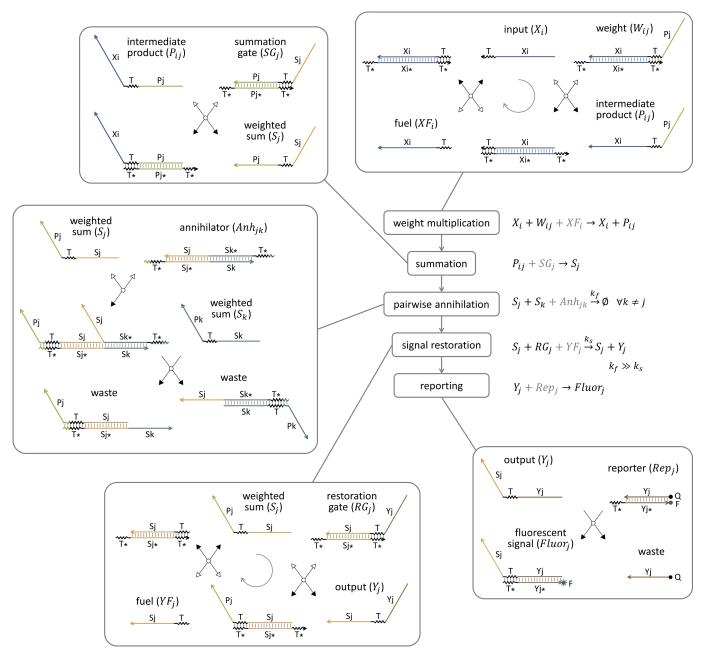
Theorem 5. If x is a pattern with c-bit deviation from a memory \overline{x}^{α} , where c < b(1-o), b is the total number of 1s in \hat{x}^{α} and o is the maximum overlap in \overline{x}^{α} and \overline{x}^{j} for all $j \neq \alpha$, and if all flipped 1s are no more than average in \overline{x}^{α} and all flipped 0s are no more than average in \overline{x}^{j} for all $j \neq \alpha$, then the winner-take-all neural network recognizes x as \hat{x}^{α} .

Proof. Let c_0 be the number of flipped 0s (that is, where 1 in \mathbf{x} and 0 in $\hat{\mathbf{x}}^{\alpha}$ appear at the same index) and c_1 be the number of flipped 1s (that is, where 0 in \mathbf{x} and 1 in $\hat{\mathbf{x}}^{\alpha}$ appear at the same index). First, $s_{\alpha} = \overline{\mathbf{x}}^{\alpha} \cdot \mathbf{x} \ge 1 - c_1/b$. Second, for $j \ne \alpha$, $s_j = \overline{\mathbf{x}}^j \cdot \mathbf{x} \ge 0 + c_0/b$. Third, because $c = c_0 + c_1$ and c < b(1 - o), $o + c_0/b = o + (c - c_1)/b < o + [b(1 - o) - c_1]/b = 1 - c_1/b$. Putting the three constraints together, we conclude that $s_{\alpha} > s_j$ and thus $y_{\alpha} = 1$ and $y_j = 0$ for all $j \ne \alpha$.

These are not the strongest results possible, but they provide intuition about how the winner-take-all neural network functions, with both binary and analogue weights, and how tolerant to errors it is.

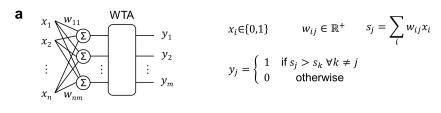
Data availability. All data that support the findings of this study are included in the manuscript and its Extended Data. Source Data for Figs. 2–4 and Extended Data Figs. 3–7 are provided with the online version of the paper.

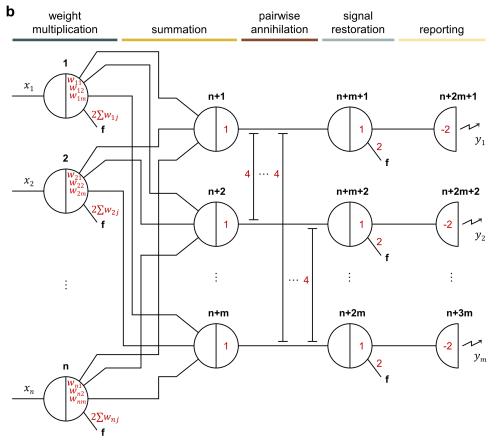
31. Zadeh, J. N. et al. NUPACK: analysis and design of nucleic acid systems. J. Comput. Chem. **32**, 170–173 (2011).



Extended Data Fig. 1 | **DNA implementation of winner-take-all neural networks.** The winner-take-all computation is broken into five subfunctions: weight multiplication, summation, pairwise annihilation, signal restoration and reporting. In the chemical reactions listed next to the five subfunctions, the species in black are needed as part of the function, the species in grey are needed to facilitate the reactions and the waste species are not shown. $k_{\rm f}$ and $k_{\rm s}$ are the rate constants of the pairwise-annihilation and signal-restoration reactions, respectively. In the DNA-strand-displacement implementation, weight multiplication and signal restoration are both catalytic reactions. The grey circle with an

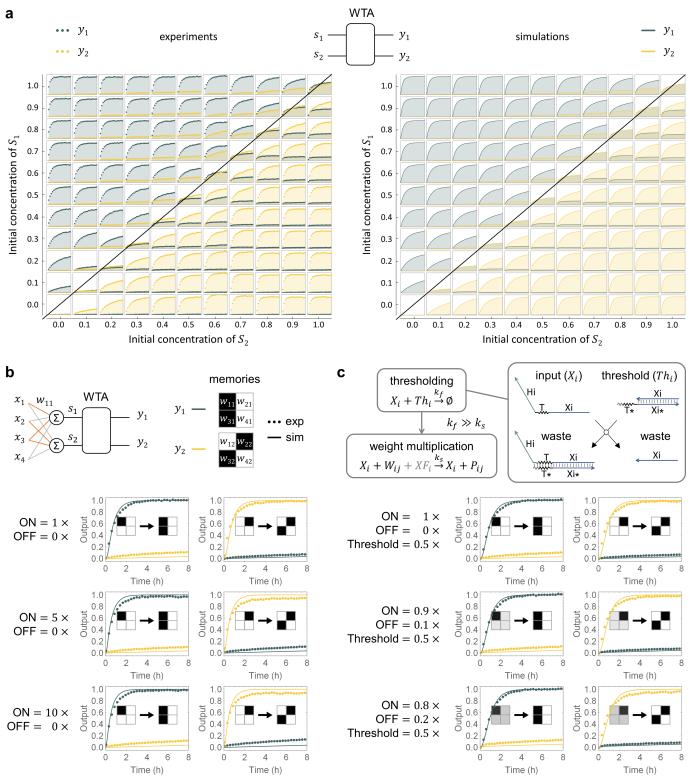
arrow indicates the direction of the catalytic cycle. Representative, but not all possible, states are shown for the pairwise-annihilation reaction. Zigzag lines indicate short (5 or 7 nucleotide) toehold domains and straight lines indicate long (15 or 20 nucleotide) branch-migration domains in DNA strands, with arrowheads marking their 3′ ends. Each domain is labelled with a name, and asterisks in the names indicate sequence complementarity. Black-filled and white-filled arrowheads indicate the forwards and backwards directions of a reaction step, respectively. All DNA sequences are listed in Supplementary Table 1.





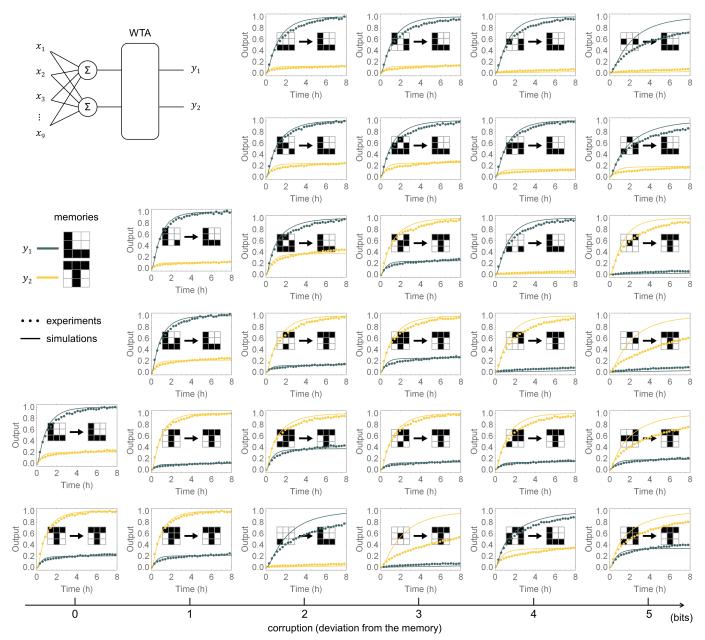
Extended Data Fig. 2 | Seesaw circuit implementation of winner-take-all neural networks. a, Same as Fig. 1a. b, Seesaw circuit diagram¹¹ for implementing the winner-take-all neural network. Each black number indicates the identity of a seesaw node. A total of n + 3m nodes are required for implementing a winner-take-all neural network with m memories that each has n bits. The location and absolute value of each red number indicates the identity and relative initial concentration of a

DNA species, respectively. A red number on a wire connected to a node (or between two nodes) indicates a free signal molecule, which can be an input or fuel strand. A red number inside a node indicates a gate molecule, which can be a weight, summation gate or restoration gate. A red number on a wire that stops perpendicularly at two wires indicates an annihilator molecule. A negative red number inside a half node with a zigzag arrow indicates a reporter molecule.



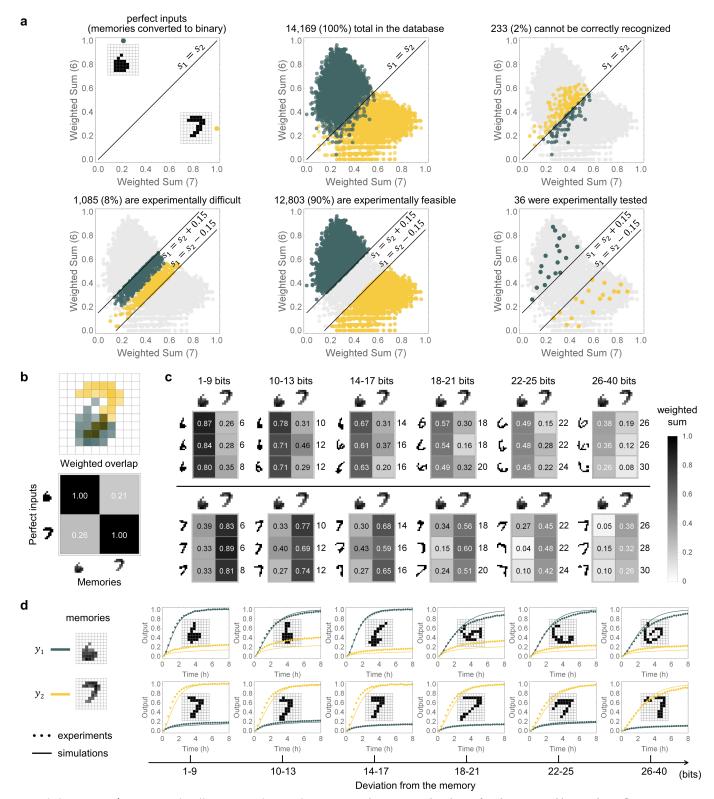
Extended Data Fig. 3 | **Experimental characterization of winner-take-all DNA neural networks.** a, Two-species winner-take-all behaviour. The experimental data (left, same as Fig. 2a) were used to identify the reverse rate constant $k_r = 0.4 \text{ s}^{-1}$ of the annihilation reaction in simulations (right). All fluorescence kinetics data and simulation are shown over the course of 2.5 h. The standard concentration is 50 nM (1×). Initial concentrations of the annihilator, restoration gates, fuels and reporters are 75 nM (1.5×), 50 nM (1×), 100 nM (2×) and 100 nM (2×), respectively. b, A 4-bit pattern recognition circuit with input concentration varying from 50 nM to 500 nM. In each output trajectory plot, dotted lines indicate fluorescence kinetics data and solid lines indicate simulation. The patterns to the left and right of the arrow indicate input signal and

output classification, respectively. **c**, Applying thresholding to clean up noisy input signals. The thresholding mechanism has been reported previously in work on seesaw DNA circuits 11 . The extended toehold in threshold molecule has 7 nucleotides. In **b** and **c**, to compare the range of inputs, the concentration of each input strand is shown relative to 50 nM. The initial concentration of each weight molecule is either 0 or 50 nM; weight fuels are twice the concentration of weight molecules. The initial concentrations of the summation gates, annihilator, restoration gates, restoration fuels and reporters are 100 nM (1×), 400 nM (4×), 100 nM (1×), 200 nM (2×) and 200 nM (2×), respectively, with a standard concentration of 100 nM.



Extended Data Fig. 4 | A winner-take-all DNA neural network that recognizes 9-bit patterns as 'L' or 'T'. In each output trajectory plot, dotted lines indicate fluorescence kinetics data and solid lines indicate simulation. The standard concentration is 50 nM (1×). The initial concentration of each input strand is either 0 or 50 nM (1×). The initial concentration of each weight molecule is either 0 or 10 nM (0.2×); weight fuels are twice the concentration of weight molecules. The initial concentrations of the summation gates, annihilator, restoration gates,

restoration fuels and reporters are 50 nM (1×), 75 nM (1.5×), 50 nM (1×), 100 nM (2×) and 100 nM (2×), respectively. The patterns to the left and right of the arrow indicate input signal and output classification, respectively. In addition to the perfect inputs, 28 example input patterns with 1–5 corrupted bits were tested. Note that 5 is the maximum number of corrupted bits, because an 'L' with more than 5-bit corruption will be as similar as or more similar to a 'T', and vice versa.

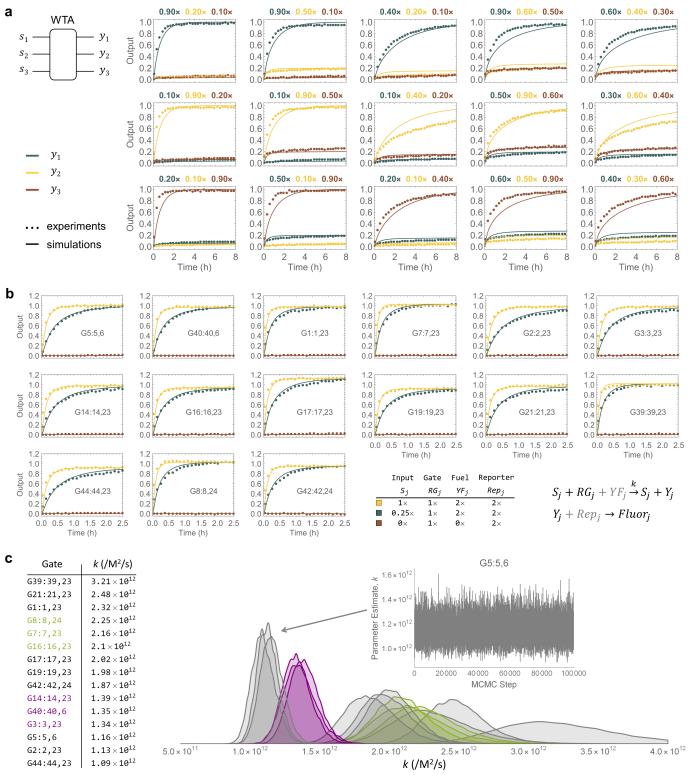


Extended Data Fig. 5 | A winner-take-all DNA neural network that recognizes 100-bit patterns as one of two handwritten digits.

a, Choosing the test input patterns on the basis of their locations in the weighted-sum space. b, Overlap between the two memories: '6' and '7'.

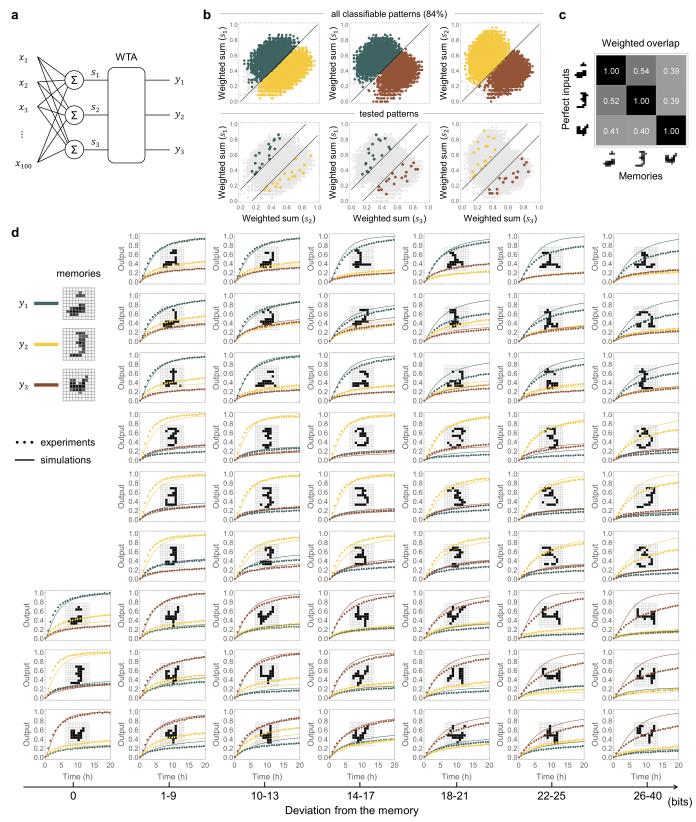
c, 36 test patterns with the number of flipped bits shown next to their weighted sums. d, Recognizing handwritten digits with up to 30 flipped

bits compared to the perfect digits. Dotted lines indicate fluorescence kinetics data and solid lines indicate simulation. The standard concentration is 100 nM. Initial concentrations of all species are listed in Extended Data Fig. 10. The input pattern is shown in each plot. Note that 40 is the maximum number of flipped bits because all patterns have exactly 20 1s.



Extended Data Fig. 6 | Three-species winner-take-all behaviour and rate measurements for selecting DNA sequences in winner-take-all reaction pathways. a, Fluorescence kinetics data for a three-species winner-take-all circuit. Initial concentrations of the three weighted-sum species are shown on top of each plot as a number relative to a standard concentration of 50 nM (1×). The initial concentrations of the annihilator, restoration gates, fuels and reporters are 75 nM (1.5×), 50 nM (1×), 100 nM (2×) and 100 nM (2×), respectively. b, Measuring the rates of 15 catalytic gates. Fluorescence kinetics data (dotted lines) and simulations (solid lines) of the signal restoration reaction are shown, with a trimolecular rate constant

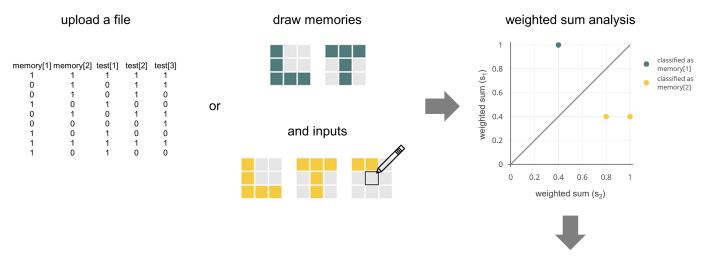
(k) fitted using a Markov chain Monte Carlo package (https://github.com/joshburkart/mathematica-mcmc). The reporting reaction was needed for the fluorescence readout. Initial concentrations of all species are listed as a number relative to a standard concentration of 50 nM. c, The 15 catalytic gates sorted and grouped on the basis of their rate constants. All rate constants are within $\pm 65\%$ of the median. The two coloured groups of three rate constants are within $\pm 5\%$ of the median. These two groups of catalytic gates were selected for signal restoration in the winner-takeall DNA neural networks that remember two to nine 100-bit patterns (Methods section 'Sequence design').



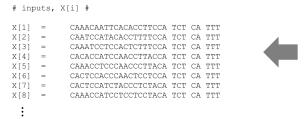
Extended Data Fig. 7 | A winner-take-all DNA neural network that recognizes 100-bit patterns as one of three handwritten digits. a, Circuit diagram. b, Choosing the test input patterns on the basis of their locations in the weighted-sum space. c, Overlap between the three memories: '2', '3' and '4'. d, Recognizing handwritten digits with up to 28 flipped bits compared

to the 'remembered' digits. Dotted lines indicate fluorescence kinetics data and solid lines indicate simulation. The standard concentration is 100 nM. Initial concentrations of all species are listed in Extended Data Fig. 10. The input pattern is shown in each plot. Note that 40 is the maximum number of flipped bits because all patterns have exactly 20 1s.





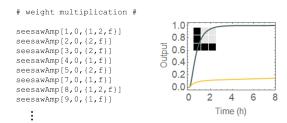
generate DNA sequences



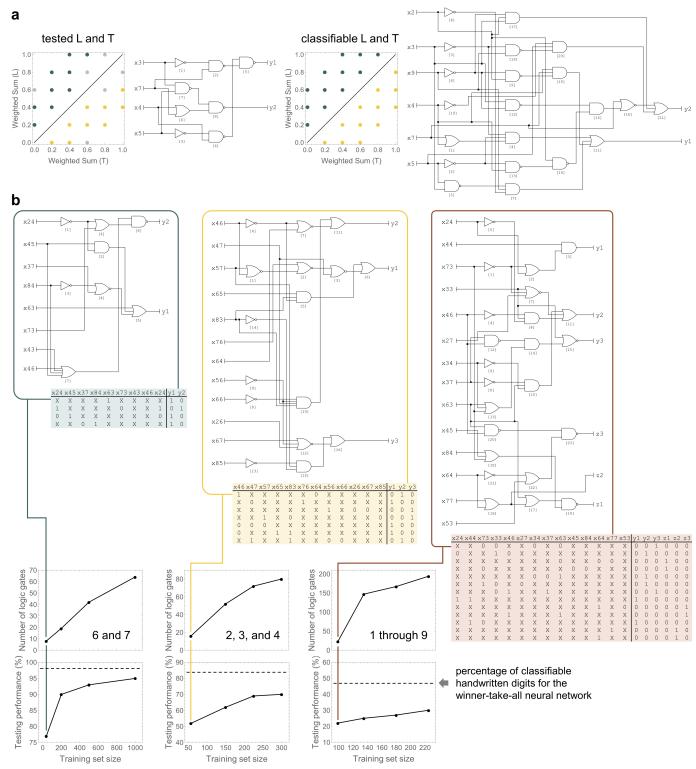
Extended Data Fig. 8 | Workflow of the winner-take-all compiler.

The compiler²¹ is a software tool for designing DNA-based winner-take-all neural networks. Users start by uploading a file that describes a winner-take-all neural network. Alternatively, the weight matrix and test patterns can be drawn graphically. Next, a plot of the weighted-sum space provides a visual representation of the classification decision boundaries.

simulate network behavior



The kinetics of the system can be simulated using Mathematica code downloaded from the compiler website, and the set of reaction functions are displayed online. Finally, the compiler produces a list of DNA strands that are required to experimentally demonstrate the network as designed by the user.



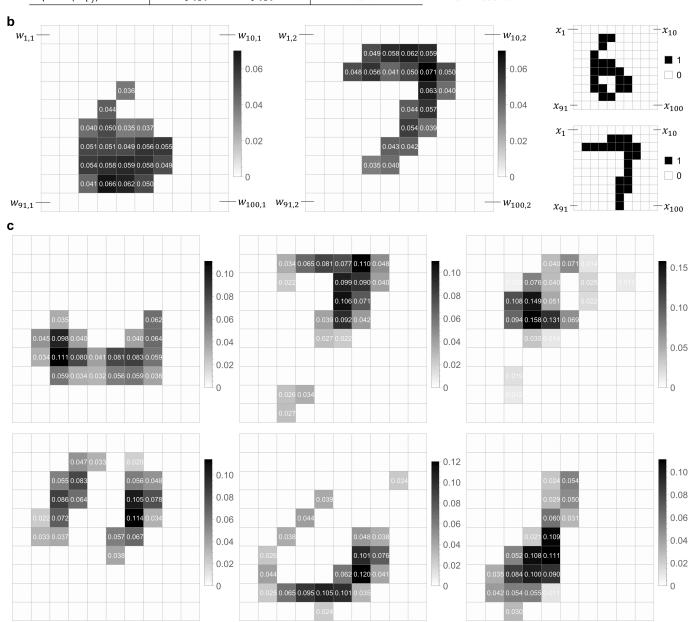
Extended Data Fig. 9 | Size and performance analysis of logic circuits for pattern recognition. a, Logic circuits that determine whether a 9-bit pattern is more similar to 'L' or 'T'. b, Logic circuits that recognize 100-bit handwritten digits. To find a logic circuit that produces correct outputs for a given set of inputs, with no constraint on other inputs, we first created a truth table including all experimentally tested inputs and their corresponding outputs. The outputs for all other inputs were specified as 'don't care', meaning the values could be 0 or 1. The truth table was converted to a Boolean expression and minimized in Mathematica, and then minimized again jointly for multiple outputs and mapped to a logic circuit in Logic Friday (https://download.cnet.

com/Logic-Friday/3000-20415_4-75848245.html). In the minimized truth tables shown here, 'X' indicates a specific bit of the input on which the output does not depend. For comparison, minimized logic circuits were also generated from training sets with a varying total number of random examples from the MNIST database. The performance of each logic circuit, defined as the percentage of correctly classified inputs, was computed using all examples in the database. To make the minimization and mapping to logic gates computable in Logic Friday, the size of the input was restricted to the 16 most significant bits, determined on the basis of the weight matrix of the neural networks.

RESEARCH LETTER

Name of species	Name of strands	Initial concentration
input (X_i)	X[i]	$x_i/20 \times$
weight (W_{ij})	P[i,j] + W[i]-b	$w_{ij} \times$
fuel (XF_i)	XF[i]	$2\sum_{j=1}^{m}w_{ij}\times$
summation gate (SG_j)	S[j] + SG[j]-b	1 ×
annihilator (Anh _{jk})	Anh[j,k]-t + Anh[j,k]-b	4 ×
restoration gate (RG _j)	Y[j] + RG[j]-b	1 ×
fuel (YF _j)	YF[j]	2 ×
reporter (Rep _i)	Rep[j]-t + Rep[j]-b	2 ×

 $1 \times = 100 \text{ nM}$



Extended Data Fig. 10 | Species and their initial concentrations in all neural networks that recognize 100-bit patterns. a, List of species and strands. Reporters were annealed with top strands (that is, Rep[j]-t) in 20% excess. All other two-stranded complexes were annealed with a 1:1 ratio of the two strands and then PAGE-purified (Methods section 'Purification').

b, Weights and example inputs in the neural network that recognizes '6' and '7'. **c**, Weights in the neural network that recognizes '1'-'9'. Weights and inputs used in all experiments are listed in Supplementary Table 2. Detailed protocols for all experiments are listed in Supplementary Table 3.



Controlling an organic synthesis robot with machine learning to search for new reactivity

Jarosław M. Granda¹, Liva Donina¹, Vincenza Dragone¹, De-Liang Long¹ & Leroy Cronin¹*

The discovery of chemical reactions is an inherently unpredictable and time-consuming process¹. An attractive alternative is to predict reactivity, although relevant approaches, such as computer-aided reaction design, are still in their infancy². Reaction prediction based on high-level quantum chemical methods is complex³, even for simple molecules. Although machine learning is powerful for data analysis^{4,5}, its applications in chemistry are still being developed⁶. Inspired by strategies based on chemists' intuition, we propose that a reaction system controlled by a machine learning algorithm may be able to explore the space of chemical reactions quickly, especially if trained by an expert8. Here we present an organic synthesis robot that can perform chemical reactions and analysis faster than they can be performed manually, as well as predict the reactivity of possible reagent combinations after conducting a small number of experiments, thus effectively navigating chemical reaction space. By using machine learning for decision making, enabled by binary encoding of the chemical inputs, the reactions can be assessed in real time using nuclear magnetic resonance and infrared spectroscopy. The machine learning system was able to predict the reactivity of about 1,000 reaction combinations with accuracy greater than 80 per cent after considering the outcomes of slightly over 10 per cent of the dataset. This approach was also used to calculate the reactivity of published datasets. Further, by using real-time data from our robot, these predictions were followed up manually by a chemist, leading to the discovery of four reactions.

Recent progress in automated chemistry^{9,10}, online analytics¹¹ and real-time optimization¹² suggests that it is possible to construct a robot that can autonomously explore chemical reactivity. With this in mind, we have designed, built and programmed a bespoke chemicalhandling robot comprising in-line spectroscopy, real-time data analysis and feedback mechanisms (Fig. 1a, b). The robot is configured to execute six experiments in parallel, allowing up to 36 experiments to be performed per day. To evaluate the outcome of a reaction, the robot is equipped with real-time sensors—a flow benchtop nuclear magnetic resonance (NMR) system¹³, a mass spectrometer and an attenuated total-reflection infrared spectroscopy system¹⁴—to record the spectra of the reaction mixtures. Then, it uses an algorithm to automatically classify the reaction mixtures as reactive or non-reactive, which is reported in binary form as zero or one, using a supported vector machine¹⁵ (SVM) with a linear kernel (Fig. 1c) model. This algorithm compares the spectrum of the starting materials with that recorded by the robotic platform using NMR and infrared spectroscopy, registering differences as reactivity hits (see Fig. 1e for an example of a non-reactive mixture and Fig. 1f for a reactive mixture). By training the model on 72 reactive and non-reactive mixtures manually classified by an expert chemist, the model could classify the reactivity of reaction mixtures with an accuracy of 86%, as determined by leave-one-out cross-validation. The machine learning algorithm used to explore the chemical space needs an automatically generated representation of the reactions¹⁶. Because the representation of the data is crucial for machine learning 17, we created a reaction descriptor with a width corresponding to the number of starting materials in the pool of reagents and with bits representing reagents that were present in a given reaction mixture to one, similarly to one-hot encoding. Figure 1d shows example vector representations for the model substrate pool consisting aniline, benzaldehyde, acetyl chloride, phenylhydrazine and furan.

This approach to representing chemical space renders it structureindependent and allows the robotic platform to operate without prior knowledge about reactivity and chemical structure (Fig. 2). Initially, the chemical space was sampled by performing reactions with random combinations of starting materials, evaluating their reactivity as reactive or non-reactive using the SVM model (to determine expected values of reactivity, Y) and encoding them in vector form (to obtain a training set, X). The process of random selection is important because the system avoids making prior assumptions about the possible reactivity of the reagents, ensuring that the initial run results are unknown. Even if the reaction mixture decomposes or is non-reactive, this information is still useful for the navigation of the chemical space, allowing real-time assessment of the reactivity of the starting materials. After the reaction database has been built, a linear discriminant analysis (LDA)¹⁸ model is trained on the data obtained to construct a model of the chemical space. The remaining reactions are then rated by predicting the probability of reactivity using the LDA model. This allows for autonomous decision making, and the reaction with the highest score is performed and analysed by the robotic system, thus avoiding many non-reactive combinations and speeding up the search. The loop is closed by updating the reaction database with the result of the last experiment from the platform and then by retraining the LDA model of the chemical space. The cycle is repeated until the required number of reactions is performed or until the whole space—defined by a pool of 18 reactive, structurally diverse molecules containing functional groups 1-18 (Extended Data Fig. 1)—is spanned. The chemical space constituted of two- and three-component reactions formed from the pool of starting materials, giving 969 possible experiments. When LDA was performed, the algorithm was able to clearly differentiate between reactive and non-reactive combinations of the starting materials (Fig. 3a). This means that the LDA can be useful for predicting new reactivity. By taking this approach, we showed that the robot can learn how reactive the starting materials are and efficiently navigate chemical space. For example, the reaction mixture composed from 2-aminothiazole (9), phenylacetyl chloride (15) and DBU (13) would be classified as highly reactive, a mixture of malononitrile (3), methylacetoacetate (18) and DBU (13) as moderately reactive and a mixture of nitromethane (4), benzofuroxan (7) and toluenesulfonylmethyl isocyanide (17) as non-reactive. These assignments agree with basic chemical intuition, demonstrating the predictive power of the model (see Supplementary Information for the reactivity of all reactions according to the LDA

To further test the learning ability of our robotic system, we performed simulations to calculate the number of reactive versus non-reactive combinations of the starting materials chosen by the algorithm during the exploration of the chemical space (Fig. 3b). In the initial stage, the space was randomly sampled, resulting in an equal number

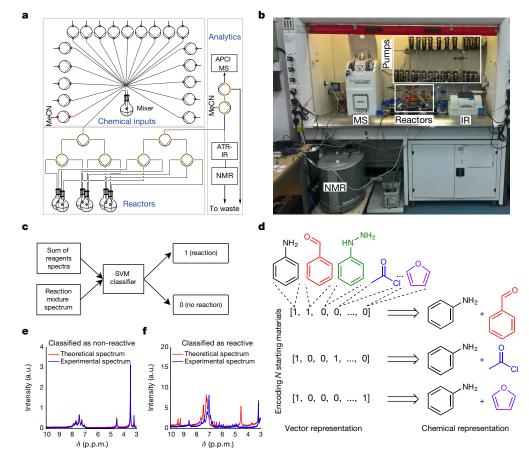


Fig. 1 | **Automatic reaction detection with machine learning. a**, Schematic of the chemical robot. The circles are pumps and the coloured dots are the positions of the valves. APCI, atmospheric pressure chemical ionization; MS, mass spectrometer; ATR-IR, attenuated total reflectance infrared spectrometer. **b**, Photograph of the chemical robot, showing the pumps, reactors and real-time analytics, including the NMR, MS

and infrared (IR) spectroscopy systems. **c**, SVM workflow for reaction detection using infrared and NMR spectroscopy, utilizing changes in the spectra. **d**, Reaction space representation using vectors. **e**, Example of a ¹H NMR (43 MHz, MeCN) spectrum for a non-reactive reaction mixture. a.u., arbitrary units. **f**, Example of a reaction mixture ¹H NMR (43 MHz, MeCN) spectrum for which a chemical reaction has been detected.

of reactive and non-reactive combinations being chosen by the algorithm. After reaching the desired number of reactions, decisions were made using LDA, leading to a rapid increase in the number of

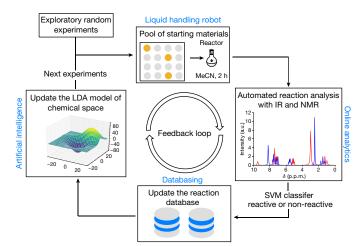


Fig. 2 | Overview of the artificial intelligence algorithm used for the exploration of chemical space with the liquid-handling robot. The liquid-handling robot performs reactions by choosing reactants from the pool of starting materials. Online analytics is used for real-time interpretation of reaction outcomes as reactive or non-reactive, and the reaction database stores reaction outcomes. Machine learning is used to build a model of the chemical space, recommend the next experiments and control the robot.

reactive combinations being chosen by the algorithm. In the end, the algorithm identified the empty part of chemical space; that is, the last experiments that were chosen were non-reactive (Fig. 3b). The accuracy of predicting the reactivity is shown in Fig. 3c, which shows that as chemical space is progressively searched, the accuracy of the prediction of the reactivity increases along with the confidence intervals. This demonstrates that the robot can 'self-learn' using artificial intelligence by exploiting this reactivity-first approach. Additionally, the accuracy of the LDA classifier in predicting the reactivity of the reaction mixtures was determined as $86\% \pm 3\%$ using five-fold cross-validation.

To further explore the predictive power of our approach, we also investigated the Suzuki-Miyaura reaction space (see Fig. 4a) described recently¹⁹ by searching for reactions with the highest yield with our machine learning approach. To achieve this, we built a neural network (for details and implementation, see Supplementary Information) and used one-hot encoding to encode literature data for machine learning. We then used the neural network to explore the hypothesis that machine learning can be used for the prediction of yields. The dataset was partitioned into a training set (3,456 reactions), a validation set (576 reactions) and a test set (1,728 reactions) to train and validate the neural network. When the neural network was tested, it performed well, giving yields with a root-mean-square error of 11% for 1,728 reactions (see Fig. 4b for the correlation between real and predicted yield). Having established that our approach can predict the yields of Suzuki-Miyaura reactions, we performed a simulation to explore this chemical space, as described above for our robot. Initially, the algorithm randomly chose 10% of the reaction space (576 reactions) and then

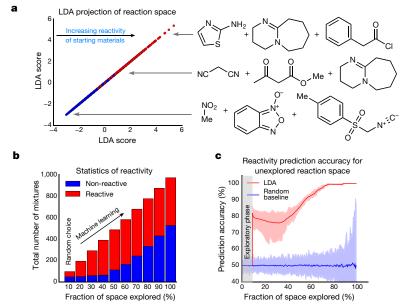


Fig. 3 | Simulations exploring the chemical space and predictive power of the model. a, Left, LDA projection of all the reactions performed, demonstrating the predictive power of LDA in classifying the reactivity. Red symbols, reactive combinations; blue symbols, non-reactive combinations. Right, examples of reactions in different regions of chemical space projected by LDA on the basis of collective chemical knowledge acquired by the robot. Top, very reactive; middle, moderately reactive;

bottom, non-reactive. **b**, Simulation showing the number of reactive and non-reactive mixtures chosen by the algorithm during the exploration of chemical space. **c**, Aggregated results from 100 simulations showing the average accuracy of the LDA in predicting the reactivity versus the fraction of chemical space explored; the confidence intervals are defined by the maximum and minium values.

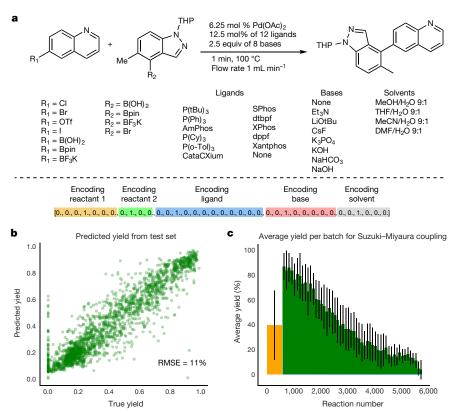


Fig. 4 | Exploring the Suzuki–Miyaura reaction using machine learning. a, The reaction space of the Suzuki–Miyaura reaction. Shown are the identity of reactants, ligand, base and solvent, and the vector representation of the reaction for machine learning. b, Validation of the predictive power of the model for a test set of 30% of the reactions (1,728 reactions). RMSE, root-mean-square error. c, Simulation of the

machine-learning-controlled exploration of this reaction space. The yellow bar shows the initial random choice of 10% of reaction space (576 reactions). The green bars show the next batches of 100 reactions chosen by the machine learning algorithm. The error bars represent the standard deviation within individual batches for Suzuki–Miyaura coupling.

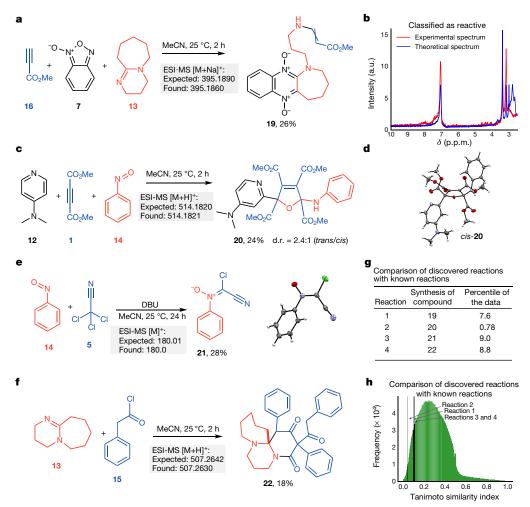


Fig. 5 | Reactivity discovered with the machine-learning-driven robot. a, Multicomponent reactions between methyl propiolate (16), benzofuroxan (7) and DBU (13); the yield obtained is given in per cent. Light-grey boxes show calculated and measured (by electrospray ionization mass spectroscopy, ESI-MS) molecular ion masses. b, ¹H NMR spectrum recorded in the platform for the reaction shown in a. c, Multicomponent reaction of DMAP (12), DMAD (1) and nitrobenzene (14), leading to the derivative 2,5-dihydrofuran (20). d.r., diastereometric

ratio. **d**, Solid-state structure of compound *cis-***20** (50% probability level). **e**, Synthesis of chlorocyanonitrone (**21**) from nitrosobenzene (**14**) and trichloroacetonitrile (**5**) in the presence of DBU (**13**). **f**, Newly discovered reaction of phenylketene with DBU. **g**, Tanimoto similarity between discovered reactions and 3.5 million known reactions. **h**, Histogram showing the Tanimoto similarity index between the discovered reactions and 3.5 million known reactions.

the neural network was trained on these data. The unexplored parts of the reaction space were then rated by the machine learning model, the next batch of candidates with the best scores was selected, and the true yield was evaluated. The initial random guess had a mean yield of 39% and standard deviation (s.d.) of 27%, shown as a yellow bar in Fig. 4c. The green bars show subsequent batches of 100 reactions chosen by the machine learning algorithm. For example, the first batch of 100 reactions had a mean yield of 85% and s.d. of 14%. The subsequent batches contained progressively fewer reactive starting materials, ultimately reaching non-reactive parts of the reaction space. This approach is valuable because it shows that by realizing only 10% of the total number of reactions, we can predict the outcomes of the remaining 90% without needing to carry out the experiments. Recently, the application of machine learning to yield prediction and the navigation of reaction space has been demonstrated for a Buchwald-Hartwig amination²⁰ and deoxyfluorination with sulfonyl fluorides²¹, leading to similar conclusions.

We used the reactive combinations discovered by the system to manually carry out reactions. For example, by analysing the spectra recorded by the robot, we identified several transformations (Fig. 5). For instance, analysis of the ¹H NMR spectrum for the reaction of methyl propiolate (**16**) with benzofuroxan (7) and DBU (**13**) suggests an interesting transformation with new peaks visible in the chemical

shift range δ = 4.0–5.0 p.p.m. and 7.9–8.5 p.p.m. (Fig. 5b). Isolation and NMR analysis of the reaction product showed that it contained protons originating from all starting materials suggesting that the compound resulted from a multicomponent reaction. Analysis of the $^1\mathrm{H}-^{^13}\mathrm{C}$ heteronuclear single-quantum and multiple-bond correlation spectra determined the structure of product **19** (see Extended Data Fig. 2a for a proposed mechanism).

We explored the utility of this reaction by synthesizing a small library of related molecules. By using substituted alkynes, we were able to prepare six structurally diverse compounds in one step (Extended Data Fig. 2b). Reaction of DMAD (1), nitrosobenzene (14) and DMAP (12) led to a multicomponent reaction with formation of 2,5dihydrofurane derivative 20 at a diastereometric ratio of 2.4:1 (trans:cis) (Fig. 5c, d). Figure 5e shows the formation of chlorocyanonitrone 21 an unreported class of nitrones—which was isolated as the product of the reaction between trichloroacetonitrile (5) and nitrosobenzene (14) in the presence of DBU (13) (structure of 21 confirmed by X-ray analysis). Finally, we also found reactivity between ketenes and DBU (Fig. 5f), indicated by the peaks at high molecular weight recorded by the platform for this reaction (mass-to-charge ratio, m/z = 506.9and m/z = 657); see Fig. 5f. Under basic conditions, phenylacetyl chloride (15) is deprotonated by DBU, giving phenyl ketene, which reacts with DBU to give the polycyclic azepine derivative 22 (Fig. 5f).

The suggested mechanisms for these transformations are presented in Extended Data Fig. 2c, d.

To assess how unique these reactions are, we employed the Tanimoto similarity index, which compares starting materials and products²². We considered over 40 million reactions, filtered by first excluding nonorganic reactions, then requiring the same number of reagents and products as our discoveries, and finally by requiring that the reactions have all the necessary structural information. This filtering left more than about 3.5 million reactions to compare. For each reaction, we calculated the similarity between each reagent and the product and calculated the mean from the obtained values. For reactions in which the reagents undergo a slight modification to reach the product, this reaction similarity index would be close to 1. Conversely, if the reagents change substantially so that the product is very different, then the result would be close to 0. All four of the reactions discovered here (see Supplementary Information) have a lower similarity index than the mean. In fact, all are in the top 10 percentile, with reaction 2 (which gives product 20) in the top 0.8 percentile (Fig. 5g), and they are considerably more distinct from the reactions chosen at random. The histogram in Fig. 5h shows that there is only one peak in the distribution and that the mean value of the Tanimoto similarity index is 0.29.

This study represents an important step towards developing intelligent automated approaches to chemical discovery using artificial-intelligence-driven chemical robots trained by human experts from the bottom up, in contrast to top-down fragment-based approaches²³.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0307-8.

Received: 18 December 2017; Accepted: 21 May 2018; Published online 18 July 2018.

- Collins, K. D., Gensch, T. & Glorius, F. Contemporary screening approaches to reaction discovery and development. Nat. Chem. 6, 859–871 (2014).
- Warr, W. A. A short review of chemical reaction database systems, computeraided synthesis design, reaction prediction and synthetic feasibility. *Mol. Inform.* 33, 469–476 (2014).
- Plata, R. E. & Singleton, D. A. A case study of the mechanism of alcoholmediated Morita Baylis-Hillman reactions. The importance of experimental observations. J. Am. Chem. Soc. 137, 3811–3826 (2015).
- 4. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. Science 349, 255–260 (2015).
- Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. Nature 533, 73–76 (2016).
- Graulich, N., Hopf, H. & Schreiner, P. R. Heuristic thinking makes a chemist smart. Chem. Soc. Rev. 39, 1503–1512 (2010).
- Gil, Y., Greaves, M., Hendler, J. & Hirsh, H. Amplify scientific discovery with artificial intelligence. Science 346, 171–172 (2014).
- Trobe, M. & Burke, M. D. The molecular industrial revolution: automated synthesis of small molecules. Angew. Chem. Int. Ed. 57, 4192–4214 (2018).

- 10. Ley, S. V., Fitzpatrick, D. E., Ingham, R. J. & Myers, R. M. Organic synthesis: march of the machines. *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).
- Sans, V. & Cronin, L. Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation. Chem. Soc. Rev. 45, 2032–2043 (2016).
- Houben, C. & Lapkin, A. A. Automatic discovery and optimization of chemical processes. Curr. Opin. Chem. Eng. 9, 1–7 (2015).
- Sans, V., Porwol, L., Dragone, V. & Cronin, L. A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy. *Chem. Sci.* 6, 1258–1264 (2015).
- Dragone, V., Sans, V., Henson, A. B., Granda, J. M. & Cronin, L. An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* 8, 15733 (2017).
- Cortes, C. & Vapnik, V. Support vector networks. *Mach. Learn.* 20, 273–297 (1995).
- Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent. Sci. 4, 268–276 (2018).
- Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal.* 35, 1798–1828 (2013).
- Coomans, D., Jonckheer, M., Massart, D. L., Broeckaert, I. & Blockx, P. Application of linear discriminant analysis in the diagnosis of thyroid diseases. *Anal. Chim.* Acta 103, 409–415 (1978).
- Perera, D. et al. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. Science 359, 429–434 (2018).
- Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C-N cross-coupling using machine learning. *Science* 360, 186–190 (2018).
- Nielsen, M. K., Ahneman, D. T., Riera, O. & Doyle, A. G. Deoxyfluorination with sulfonyl fluorides: navigating reaction space with machine learning. J. Am. Chem. Soc. 140, 5004–5008 (2018).
- Bajusz, D., Racz, A. & Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* 7, 20 (2015).
- Palazzolo, A. M. E., Simons, C. L. W. & Burke, M. D. The natural productome. Proc. Natl Acad. Sci. 114, 5564–5566 (2017).

Acknowledgements We acknowledge financial support from the EPSRC (grants number EP/H024107/1, EP/I033459/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/K038885/1, EP/L015668/1 and EP/L023652/1) and the ERC (project 670467 SMART-POM). J.M.G. acknowledges financial support from the Polish Ministry of Science and Higher Education grant number 1295/MOB/IV/2015/0. We thank A. Henson for help with the Tanimoto analysis.

Author contributions L.C. conceived the idea, developed the initial algorithms, designed the project and coordinated the efforts of the research team. J.M.G. developed the machine learning algorithms and devised the LDA and built and programmed the chemical robot. J.M.G. conducted experiments and isolated and characterized new compounds with input from L.D. and V.D. J.M.G. and L.C. co-wrote the paper with input from all authors.

Competing interests L.C. is the founder and director of DeepMatterGroup PLC and is listed as an inventor on a patent application filed by The University of Glasgow (GB 1810944.7).

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0307-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0307-8.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.C. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



METHODS

General experimental remarks. Reagents were from Sigma Aldrich and were used as received. Acetonitrile employed as a solvent in the platform was HPLC grade (VWR International). Mass spectra were recorded on a time-of-flight mass spectrometer (MicroTOF-Q MS) equipped with an electrospray source supplied by Bruker Daltonics Ltd. All data were collected in positive ion mode. The spectrometer was calibrated with a standard tune mix to give a precision of about 1.5 p.p.m. in the region m/z = 100-3,000. NMR data were recorded using a Bruker Avance III 600 MHz or a Bruker Avance 400 MHz NMR system. The spectra were recorded at 298 K using residual-solvent proton peaks for scale reference (for example, 1 H: δ (CDCl₃) = 7.26; 13 C: δ (CDCl₃) = 77.16). The chemical shifts are reported in p.p.m. using the δ scale and all coupling constants (J) are given in Hz. The following abbreviations are used to characterize spin multiplicities: s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet; dd, double doublet; dt, double triplet; dq, double quadruplet; and ddt, double doublets of triplets. Spectra obtained using distortionless enhancement by polarization transfer, correlation spectroscopy, heteronuclear single-quantum and multiple-bond correlation spectroscopy and rotating frame Overhauser-effect spectroscopy were used for structure determination and structural assignments. New reaction candidates were analysed using thin-layer chromatography (TLC) and visualized using TLC plates with a fluorescent indicator.

Syringe pumps and tubing. Control over the fluids was achieved using 27 pumps (model C3000, Tricontinent) equipped with 5 ml syringes (TriContinent) and a four-way solenoid valve according to the requirements of the experiments. The pumps were connected using a RS232 port and a daisy chain, allowing the connection of up to 16 pumps on a single RS232 bus. Commands to the pumps were sent using the pumps' proprietary control language, implemented in a Python module, allowing control over the pumps and error-reporting functionality (for example, pumps malfunctioning). PTFE plastic tubing with an outer diameter of 1/8 inch (3.175 mm) was cut to the specified length and connected using standard HPLC low-pressure PTFE connectors and PEEK manifolds (supplied by Kinesis). Online attenuated total-reflectance infrared spectroscopy. All spectra were recorded using a Thermo ScientificNicoletiS5 Fourier transform infrared spectroscopy system equipped with a ZnSe Golden Gate attenuated total reflectance infrared flow cell. The resolution was set at 4 cm⁻¹ and each sample's spectrum was recorded using 36 scans. The spectrometer was controlled by OMNIC software using Python and the ActiveX software framework. Before measurement of the spectra, the solvent (MeCN) was recorded as background.

Online NMR spectroscopy. The NMR spectra were recorded using a Spinsolve benchtop NMR system from Magritek with a compact permanent magnet (43 MHz) based on the Hallbach design, working on a lock-free basis (not requiring deuterated solvents). Shimming was performed using a D₂O/H₂O mixture (9:1 v:v) to minimize the half-width of the solvent peak. To measure reaction mixtures, the spectrometer was equipped with a home-built flow cell with a standard 5 mm width to maximize sensitivity. The spectra were measured in a stopped flow by pumping reaction mixtures into the flow cell. The spectrometer was controlled by Spinsolve software by sending XML messages over a network connection.

Benchtop mass spectroscopy. The spectra were recorded with an Advion Expression mass spectrometer using the atmospheric pressure chemical ionization technique. The detailed acquisition parameters can be found in Supplementary Information. The mass spectrometer was controlled using Python wrapper software and Advion API, allowing complete control over the instrument and acquisition parameters. Dilution of the reaction mixtures, which was necessary for recording their spectra, was realized using two syringe pumps by diluting reaction mixtures 3,125 times using solvent (MeCN) before the measurements.

Flow setup implementation. The platform was assembled as in Fig. 1a, using the 27 syringe pumps, the benchtop infrared spectroscopy system, the NMR and the mass spectrometer. Round bottom flasks (25 ml) were employed as the mixer and reactors. 18 pumps were responsible for dispensing the chemicals to the mixer, six pumps were used to transfer the reaction mixture from the mixer to the proper reactor, one pump was employed to pump the solvent (MeCN), and two pumps

were used to realize the dilution step that was necessary to measure mass spectra. The starting materials were prepared as 1.0 M solutions. Automatic data collection and processing and platform control were achieved using the Python programming language. Before the execution of the reaction, the robot was cleaned three times by flushing the mixer, reactor flasks and analytics. The reaction was performed by adding proper reagents to the mixer (total volume 5.0 ml) in a 1:1 ratio, transferring the reaction mixture to the reactor and saving the reaction parameters (the identity and volumes of the starting materials). After two hours, the reaction mixture was transferred to the measurement loop, where the NMR and infrared spectra were recorded. The mass spectrum was recorded after dilution of the reaction mixture. After the reaction mixture has been measured, the mixer, reactor and analytics were cleaned by flushing with solvent twice. Parallel execution of six reactions was implemented by shifting the execution of each reaction in time so that each experiment had access to the liquid-handling robot and analytics without colliding with the other experiments. Spectra (NMR and infrared) were also recorded for each chemical in the pool of starting materials (Extended Data Fig. 1) that was used for the calculation of the theoretical spectrum of the reaction mixture.

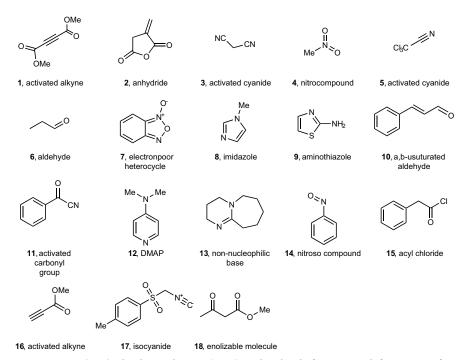
Autonomous navigation of chemical space by the robot. The algorithm for the exploration of chemical space starts by measuring 90 random experiments in the platform, and then each experiment in this set is processed to assess its reactivity and generate its vector representation. The ¹H NMR spectrum of the reaction mixture is automatically processed using fast Fourier transform, phasing and referencing of the solvent peak. The intensity of the solvent peak is normalized to 1.0 (the solvent peak is used as an internal standard, allowing easy addition of the spectra). The infrared spectra are used without any preprocessing. Next, the theoretical spectra of the reaction mixture (the sum of the starting materials) are constructed for NMR and infrared spectroscopy. The spectra are normalized by removing the mean and scaled to unit variance. The reactivity of the reaction mixture is assessed by feeding the NMR reaction mixture and NMR theoretical spectrum to the SVM classifier (trained previously; see Supplementary Information). The outcome of the classifier is Y = 0 (non-reactive) or Y = 1 (reactive). Similarly, the reactivity is assessed by the SVM classifier using the infrared spectra. An experiment is classified as reactive if any of the above classifiers categorizes it as reactive. The vector representation is generated using the identity of the starting materials. The vector representation (*X*) and reactivity (*Y*) are added to the reaction database.

The machine learning algorithms are realized using the sci-kit learn package in Python²⁴. After the initial database of the reactions is built, the LDA classifier is trained on the vector representation of the reactions (*X*) and their reactivity (*Y*). All the possible unperformed reactions are then rated by assigning them the probability of being reactive, as calculated from the LDA model. After the reactions with the highest score are realized by the liquid-handling robot, they are processed as described above, updating the reaction database. Then, the LDA model is retrained on the updated database and the robot iteratively explores the chemical space until the desired number of experiments is performed. Simulations of the exploration of the chemical space with this algorithm were performed using the data collected by the robot.

Syntheses of molecules discovered by the platform. The solutions of the starting materials (1.0 M solutions in MeCN) were added to the round bottom flask (25 ml) in a 1:1 ratio (total volume 5.0 ml) and stirred in room temperature for 2 h. Subsequently, silica gel (4.0 g) was added and the solvent was evaporated. The products of the reaction were isolated using column chromatography. The syntheses of all compounds were adjusted according to the need for each reaction. For the detailed procedure followed for each compound and characterization, see Supplementary Information.

Data and code availability. The data used for simulations of the exploration of chemical space are available in Supplementary Information. The code and data can be found online at https://github.com/croningp/reaction_learning. The data used for Suzuki–Miyaura coupling are available in ref. ¹⁹.

24. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).



Extended Data Fig. 1 | Reaction space explored. The chemical inputs (1–18) used in the platform to search for new transformations and to evaluate the performance of the algorithm.

Extended Data Fig. 2 \mid Suggested mechanisms for observed transformations and small library of compounds synthesized.

a, Suggested mechanism for the synthesis of compound 19. b, Small library

of compounds synthesized. c, Suggested mechanism for the synthesis of compound 22. d, Suggested mechanism for the synthesis of compound 21.



Synchronous tropical and polar temperature evolution in the Eocene

Margot J. Cramwinckel^{1*}, Matthew Huber², Ilja J. Kocken¹, Claudia Agnini³, Peter K. Bijl¹, Steven M. Bohaty⁴, Joost Frieling¹, Aaron Goldner², Frederik J. Hilgen¹, Elizabeth L. Kip¹, Francien Peterse¹, Robin van der Ploeg¹, Ursula Röhl⁵, Stefan Schouten^{1,6} & Appy Sluijs¹

Palaeoclimate reconstructions of periods with warm climates and high atmospheric CO2 concentrations are crucial for developing better projections of future climate change. Deep-ocean^{1,2} and high-latitude³ palaeotemperature proxies demonstrate that the Eocene epoch (56 to 34 million years ago) encompasses the warmest interval of the past 66 million years, followed by cooling towards the eventual establishment of ice caps on Antarctica. Eocene polar warmth is well established, so the main obstacle in quantifying the evolution of key climate parameters, such as global average temperature change and its polar amplification, is the lack of continuous high-quality tropical temperature reconstructions. Here we present a continuous Eocene equatorial sea surface temperature record, based on biomarker palaeothermometry applied on Atlantic Ocean sediments. We combine this record with the sparse existing data⁴⁻⁶ to construct a 26-million-year multi-proxy, multi-site stack of Eocene tropical climate evolution. We find that tropical and deep-ocean temperatures changed in parallel, under the influence of both long-term climate trends and short-lived events. This is consistent with the hypothesis that greenhouse gas forcing^{7,8}, rather than changes in ocean circulation 9,10, was the main driver of Eocene climate. Moreover, we observe a strong linear relationship between tropical and deep-ocean temperatures, which implies a constant polar amplification factor throughout the generally icefree Eocene. Quantitative comparison with fully coupled climate model simulations indicates that global average temperatures were about 29, 26, 23 and 19 degrees Celsius in the early, early middle, late middle and late Eocene, respectively, compared to the preindustrial temperature of 14.4 degrees Celsius. Finally, combining proxy- and model-based temperature estimates with available CO₂ reconstructions⁸ yields estimates of an Eocene Earth system sensitivity of 0.9 to 2.3 kelvin per watt per square metre at 68 per cent probability, consistent with the high end of previous estimates11.

It is well established that deep-ocean temperatures peaked during the Early Eocene Climatic Optimum (EECO; about 52–50 million years (Myr) ago) and had declined substantially by the latest Eocene (about 34 Myr ago)^{1,2}. These trends are mimicked in reconstructions of sea surface temperature (SST) in the southern high latitudes³ because Eocene deep-ocean temperatures reflect Southern Ocean winter surface conditions that are relayed to the abyss through deep-water formation¹². However, to unlock the unique promise of Eocene palaeoclimate records to answer fundamental questions about the relationship between atmospheric CO₂ concentrations and global temperature, and to quantify the polar amplification of climate change, accurate reconstructions of tropical surface oceans are required. Moreover, tropical records are necessary to test the two competing hypotheses for Eocene deep-ocean and polar cooling: (1) decreasing greenhouse gas concentrations, predominantly CO₂ (refs ^{7,8}), and (2) changes in ocean

circulation and meridional heat transport associated with opening of ocean gateways 9,10 . In theory, gateway opening cools the Southern Ocean and deep ocean while warming the upper tropical ocean by a few degrees 13 , whereas CO_2 decline leads to global cooling at both the Equator and the poles 14 —albeit with predicted amplified polar temperature change relative to the tropics 15 . Importantly, this amplification factor affects the volume and extent of ice sheets, and thus the global sea level, and is therefore critical to constrain, also for future projections. Yet, despite evidence for CO_2 decline over the Eocene 8 , existing tropical records $^{4-6}$ are fragmentary and of low resolution, and therefore insufficient to address these crucial questions.

We generated new temperature reconstructions using a clay-bearing, micritic porcellanite sequence recovered at Ocean Drilling Program (ODP) Site 959 in the eastern equatorial Atlantic Ocean (Fig. 1). Site 959 was positioned at near-equatorial latitudes and deep-bathyal water depths throughout the Eocene¹⁶ (Extended Data Table 1a). We augment the existing age model¹⁶ with new biostratigraphic and chemostratigraphic constraints (Extended Data Table 1b, Extended Data Fig. 1). Although carbonate preservation is poor¹⁶, well preserved, immature organic matter is present throughout¹⁷. We therefore employ the organic TEX₈₆ palaeothermometer, which utilizes the temperature-dependent distribution of thaumarchaeotal membrane lipids to reconstruct SST. Fractional abundances of the various lipids at Site 959 indicate an upper water column (50–300 m) source (Methods), which allows confident SST interpretations from TEX₈₆. Several calibrations exist to translate TEX₈₆ data into SSTs on the basis of a modern core–top

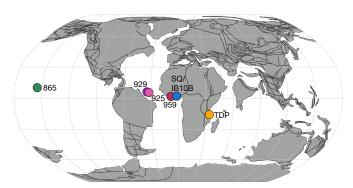


Fig. 1 | Palaeogeographic reconstruction of the studied sites 40 million years ago. The figure shows the approximate palaeoposition of the studied site (ODP Site 959) and the main sites that we used to produce a tropical SST compilation: ODP sites 865, 925 and 929; Tanzania Drilling Project (TDP); Sagamu Quarry (SQ) and IB10B Core, Nigeria. Continental plates are shown in dark grey. Light-grey gridlines represent latitudes and longitudes, with 30° spacing. The map was generated with GPlates, using the rotation frame and tectonic reconstruction of Matthews et al. 30.

¹Department of Earth Sciences, Faculty of Geoscience, Utrecht University, Utrecht, The Netherlands. ²Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA. ³Department of Geosciences, University of Padova, Padova, Italy. ⁴Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK. ⁵MARUM - Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. ⁶NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry and Utrecht University, Den Burg, The Netherlands. *e-mail: m.j.cramwinckel@uu.nl

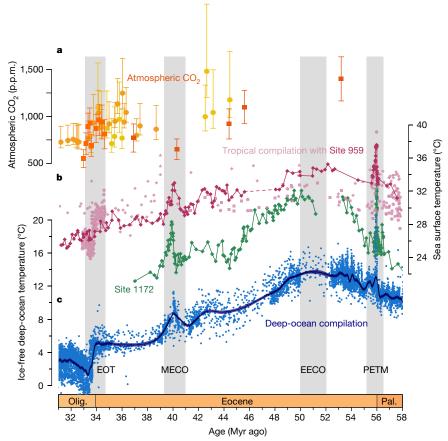


Fig. 2 | Eocene global climate evolution. a, CO_2 record from boron isotopes from the TDP (orange squares; error bars represent 68% confidence intervals) and alkenones from ODP Sites 612 and 925 (yellow and orange circles; uncertainties from original studies); data sources are provided in Methods. b, TEX_{86}^{H} -based SST record for Site 959 (red) and additional tropical compilation (pink; see Extended Data Fig. 5). The

dashed line represents a hiatus. Green diamonds (Site 1172) show a high-latitude TEX $_{86}^{\rm H}$ -based SST record³. c, $\delta^{18}{\rm O}$ -based ice-free deep-ocean temperature (described in Methods), with fitted LOESS model (black line) and 95% confidence interval (dark-blue shading). Age follows the Geologic Time Scale 2012 (GTS2012). Pal., Palaeocene; Olig., Oligocene; EOT, Eocene–Oligocene transition.

dataset¹⁸. For biophysical and analytical reasons, we prefer conservative estimates of tropical temperature generated by the logarithmic TEX_{86}^{H} calibration¹⁸ (Methods, Extended Data Fig. 2). In addition, we use a linear Bayesian spatially varying regression (BAYSPAR) calibration¹⁹ as complementary analysis (Extended Data Fig. 3).

Our new equatorial record from Site 959 (Fig. 2) shows latest-Palaeocene (about 58-56 Myr ago) SSTs of 31-33 °C, mimicking time-equivalent SSTs derived from glassy preserved planktonic foraminiferal oxygen isotope (δ^{18} O) and Mg/Ca ratios, as well as TEX $_{86}^{H}$ data from a nearby section in Nigeria²⁰—supporting the notion that TEX $_{86}^{H}$ accurately reflects SST at Site 959. The record further reveals warming by 2-3 °C from the latest Palaeocene to the earliest Eocene (58 to 53 Myr ago) to peak EECO temperatures of 34-35 °C. Superimposed transient warming of around 4°C to around 37°C occurred during the Palaeocene-Eocene Thermal Maximum (PETM), about 56 Myr ago²¹. A long-term SST drop of about 7 °C to about 28 °C characterizes the interval from the middle to late Eocene, and an additional cooling by around 2°C to around 26 °C marks the Eocene-Oligocene transition (about 34 Myr ago). Superimposed on long-term cooling is the first tropical SST reconstruction of the Middle Eocene Climatic Optimum²² (MECO), at roughly 40 Myr ago, displaying warming by around 4°C from background temperatures to a peak of about 33 °C. This provides compelling evidence that the MECO was associated with global warming; surface warming was previously only recognized in extratropical regions of the Southern Hemisphere. We also record pre-MECO temperature variability of similar duration but lower amplitude.

To assess whether regional upwelling ¹⁷ at Site 959 influenced TEX₈₆-based SST variability, we consider published total organic carbon

(TOC) contents¹⁷ of sediments and generate dinocyst assemblage data, as dinocysts are highly sensitive to upwelling in modern and Palaeogene oceans²³ (Extended Data Fig. 4). The continuous presence of cysts of Protoperidiniaceae (derived from heterotrophic dinoflagellates) and elevated TOC within biosiliceous sediments¹⁶ indicate upwelling throughout the middle and late Eocene. The early Eocene is less well constrained, but presence of Protoperidiniaceae and abundant biosilica also suggests upwelling. An upper-Eocene increase in TOC content¹⁷ might indicate upwelling intensification. Although this may exaggerate latest-Eocene cooling at Site 959, the recorded magnitude (about 2 °C) is similar to previous work at tropical locations²⁴ (Fig. 2). Apart from the late Eocene, however, variations in our SST record are not strongly correlated to changes in the abundance of upwellingindicative dinocysts or TOC content. Regional upwelling may have muted SSTs by a few degrees. Indeed, our values are somewhat lower than the few time-equivalent data points from the warm pool sampled in Tanzania⁴, suggesting that we sampled the first Eocene analogue to the 'cold tongue' in the modern ocean. Importantly, this analysis indicates that variations in the strength of upwelling were not a major factor governing SST change at the study site.

We combine our equatorial Site 959 SST record with the available low-resolution data derived from a suite of SST proxies from the Indian, Atlantic and Pacific tropical oceans (Fig. 2; data sources in Extended Data Fig. 5). Because each of these proxies is subject to different systematic sources of error, the close correspondence between various organic and carbonate proxies in both absolute temperatures and trends indicates a robust convergent temperature signal. A local regression (LOESS) model is applied to the resulting compilation to produce an

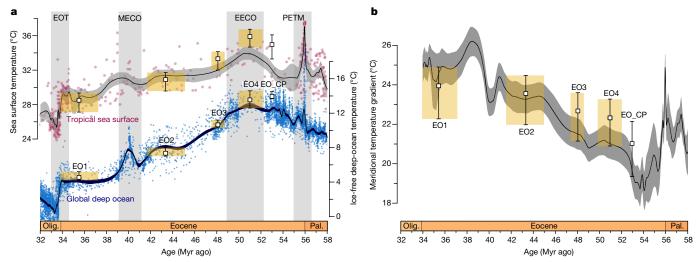


Fig. 3 | Proxy-model synthesis of Eocene temperatures. a, Top, tropical SST compilation (red) and LOESS model (black line) with 95% confidence interval (grey shading). Bottom, deep-ocean temperatures from Fig. 2c. Open squares are mean modelled tropical SSTs and deep-ocean temperatures of simulations EO1 (560 p.p.m. CO₂), EO2 (1,120 p.p.m. CO₂), EO3

(2,240 p.p.m. CO₂), EO4 (4,480 p.p.m. CO₂) and EO_CP²⁸; errors represent seasonal range. Yellow shadings illustrate age ranges to which the simulations are matched. **b**, Calculated MTG based on LOESS fits of proxy data (line, propagated 95% confidence intervals) and model simulations EO1–EO4 and EO_CP (as in **a**). Age follows GTS2012.

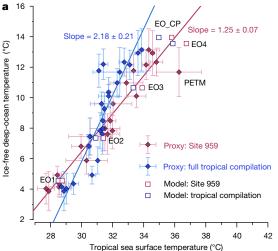
estimate of Eocene mean tropical temperature (Fig. 3a), yielding 4–7 °C of cooling through the Eocene. Remarkably, long-term trends and sub-million-year (MECO and PETM) tropical SST variations mimic those from the Southern Ocean and the deep ocean, on the basis of an updated compilation of benthic foraminiferal $\delta^{18}\text{O-derived temperatures}$ (Fig. 2). A sensitivity study indicates that potential late Eocene Antarctic ice caps did not appreciably affect this deep-ocean temperature proxy (Methods, Extended Data Fig. 6). The close correspondence between tropical and deep-sea temperatures provides solid proof that greenhouse gas forcing, rather than ocean circulation change, caused Eocene cooling, as has been suggested elsewhere 5,7,8 .

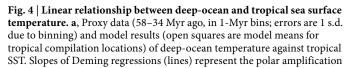
As an approximation of the pole-to-Equator temperature difference, or meridional temperature gradient (MTG), we calculate the difference between tropical mean SST and deep-ocean temperatures (Methods, Fig. 3b). Although different TEX₈₆ calibrations result in slightly different early Eocene MTGs (Extended Data Fig. 3), the gradient generally increases with cooling climate and vice versa, reflecting polar amplification of temperature variability. Remarkably, regression analysis indicates a strong linear relationship between deep-ocean and tropical temperatures (Fig. 4a; also between high-latitude and tropical SSTs in Extended Data Fig. 7). Although uncertainty on the exact value is large owing to uncertainties in temperature proxies and calibrations (Fig. 4b, Extended Data Fig. 3), this signifies a stable polar amplification factor throughout the Eocene. Because the obtained values are consistent with polar amplification derived from an analysis of the PETM event²⁰ with better spatial resolution, this observation holds for both short (10⁵-yr) and long (multi-million-year) timescales. In the absence of pronounced snow and ice albedo feedbacks, the polar amplification factor should be determined by atmospheric feedbacks²⁵. Therefore, the stable amplification factor implies that the strength of these feedbacks scales linearly with temperature in an ice-free world.

Our temperature proxy compilations provide a concrete and robust test of the ability of models to reproduce past warm climates under increased greenhouse gas forcing. We performed fully coupled general circulation model simulations using the NCAR Community Earth System Model, version 1 (CESM 1), by applying a range of radiative forcings equivalent to a range of Eocene CO₂ concentrations—560 parts per million (p.p.m.), 1,120 p.p.m., 2,240 p.p.m. and 4,480 p.p.m.; simulations EO1–EO4, respectively (see Methods)— run to full equilibrium. Because the close correspondence between tropical, high-latitude and deep-sea temperature trends (Fig. 2) supports model-based inferences that Eocene global mean temperature was relatively insensitive

to variations in palaeogeography^{7,13,26}, we did not vary the palaeogeographic boundary conditions. The modelled deep waters derive primarily from polar surface waters¹², justifying our use of the modelled and proxy-derived vertical gradient as an approximation for the MTG. The four simulations, EO1-EO4, were associated with specific age ranges by matching the simulated deep-ocean temperatures to the proxy-based deep-ocean temperatures, thus leaving SST as the predicted variable. Crucially, the simulations closely approximate the multi-proxy, multilocation tropical SST compilation for these four time slices (Fig. 3a). Therefore, the Eocene temperature gradients of 19-26 °C, which are reconstructed from proxies, are also closely reproduced (Fig. 3b). This implies that current-generation climate models are capable of resolving the low-temperature-gradient problem¹⁴ of Eocene greenhouse climates, provided sufficient greenhouse forcing, albeit with two important exceptions. First, regional proxy-model data mismatches for absolute temperatures in the South Pacific¹² and Arctic²⁷ oceans remain a conundrum, which this study does not resolve. Second, the model simulations do not fully reproduce the most reduced proxy-derived gradients of the early Eocene. On the basis of recent modelling experiments with tuned cloud parameters²⁸, one potential explanation could be that the early Eocene hothouse experienced different cloud behaviour and shortwave radiative feedbacks (simulation EO_CP in Fig. 3b). Although a simulation with tuned clouds produces a more reduced early Eocene MTG at lower CO₂ concentrations, the same parameters lead to a poorer simulation of the MTG during the PETM²⁸ (Extended Data Fig. 8), indicating that this remains an unresolved problem.

With the overall excellent agreement between ocean temperature proxy reconstructions and model simulations, we can use the latter to estimate global mean temperatures, which are required to calculate climate sensitivity to $\rm CO_2$ forcing. Global mean temperatures were about 29 °C, 26 °C, 23 °C and 19 °C during the early (54–49 Myr ago), early middle (48–46 Myr ago), late middle (42–41 Myr ago) and late Eocene (38–35 Myr ago), respectively, compared to a preindustrial temperature of 14.4 °C. These may be slightly underestimated if South Pacific³. and Arctic² temperature reconstructions represent accurate estimates of annually averaged SST. However, our model requires much larger changes in $\rm CO_2$ to produce the large and dynamic range of Eocene tropical SST and deep-sea temperature than that reconstructed from proxy data8. This implies that the Earth system sensitivity¹¹¹ to $\rm CO_2$ doubling derived from the model (3.5 °C) is too low to create sufficient warmth. We consider available Eocene $\rm CO_2$ reconstructions8 in combination with our proxy- and model-based temperatures





(Methods) to estimate the Earth system sensitivity at various parts of the Eocene (Extended Data Fig. 9). Our probabilistic analysis for the cooling between the early and late Eocene results in a calculated proxy-based Earth system sensitivity range of 0.9–2.3 K W $^{-1}$ m $^{-2}$ (68% highest density interval, equivalent to 3.5–8.9 °C per CO $_2$ doubling), consistent with the high end of previous estimates 11 .

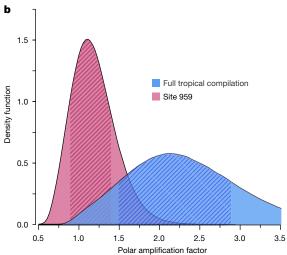
The large range of Eocene tropical temperatures on both short and long timescales indicates that the tropics respond strongly to changes in greenhouse gases, even at high temperatures. In addition to high absolute temperatures of up to about 35 °C and 37 °C during the EECO and PETM²¹, respectively, this refutes the notion of stable tropical temperatures⁴, kept constant through a physical 'thermostat' mechanism²⁹. Moreover, our results show that tropical SST varied in tandem with high-latitude and deep-ocean temperatures, with a stable Eocene polar amplification factor, consistent with a dominant role of CO₂ forcing in both long-term Eocene climate evolution and superimposed aberrations including the PETM and MECO. Tropical temperatures are expected to rise in response to anthropogenic greenhouse gas emissions. Given the consistency between our climate simulations and reconstructions, current-generation fully coupled climate models are likely to perform adequately in predicting future tropical SST change, although accurate determination of the sensitivity of global climate to CO₂ change remains a major challenge.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0272-2.

Received: 27 June 2017; Accepted: 13 April 2018; Published online 2 July 2018.

- Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. Science 287, 269–272 (2000).
- Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279–283 (2008)
- Bijl, P. K. et al. Early Palaeogene temperature evolution of the southwest Pacific Ocean. Nature 461, 776–779 (2009).
- Pearson, P. N. et al. Stable warm tropical climate through the Eocene epoch. Geology 35, 211–214 (2007).



factor (± 1 standard error). The bin 55–56 Myr for Site 959 falls to the right of the regression line because it includes a large amount of PETM event data for Site 959, but less for the deep ocean. **b**, Density function of polar amplification factor with full propagation of errors (Methods). The hatched regions show the 68% highest density intervals.

- Inglis, G. N. et al. Descent toward the Icehouse: Eocene sea surface cooling inferred from GDGT distributions. Paleoceanography 30, 1000–1010 (2015).
- Evans, D. et al. Eocene greenhouse climate revealed by coupled clumped isotope-Mg/Ca thermometry. *Proc. Natl Acad. Sci. USA* https://doi. org/10.1073/pnas.1714744115 (2018).
- Huber, M. et al. Eocene circulation of the Southern Ocean: was Antarctica kept warm by subtropical waters? *Paleoceanography* 19, PA4026 (2004).
- Anagnostou, E. et al. Changing atmospheric CO₂ concentration was the primary driver of early Cenozoic climate. Nature 533, 380–384 (2016).
- Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoceanography. J. Geophys. Res. 82, 3843–3860 (1977).
- Bijl, P. K. et al. Eocene cooling linked to early flow across the Tasmanian Gateway. Proc. Natl Acad. Sci. USA 110, 9645–9650 (2013).
- PALAEOSENS Project Members. Making sense of palaeoclimate sensitivity. Nature 491, 683–691 (2012); erratum 494, 130 (2013).
- Hollis, C. J. et al. Early Paleogene temperature history of the Southwest Pacific Ocean: reconciling proxies and models. Earth Planet. Sci. Lett. 349–350, 53–66 (2012); erratum 374, 258–259 (2013).
- Sijp, W. P., England, M. H. & Huber, M. Effect of the deepening of the Tasman Gateway on the global ocean. *Paleoceanography* 26, PA4207 (2011).
- Huber, M. & Caballero, R. The early Eocene equable climate problem revisited. Clim. Past 7, 603–633 (2011).
- Lunt, D. J. et al. A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP. Clim. Past 8, 1717–1736 (2012).
- Mascle, J. et al. Proceedings of the Ocean Drilling Program Initial Reports Vol. 159 (ODP/TAMU, College Station, 1996).
- Wagner, T. Late Cretaceous to early Quaternary organic sedimentation in the eastern Equatorial Atlantic. Palaeogeogr. Palaeoclimatol. Palaeoecol. 179, 113–147 (2002).
- Kim, J.-H. et al. New indices and calibrations derived from the distribution of crenarchaeal isoprenoid tetraether lipids: implications for past sea surface temperature reconstructions. *Geochim. Cosmochim. Acta* 74, 4639–4654 (2010).
- Tierney, J. E. & Tingley, M. P. A. Bayesian, spatially-varying calibration model for the TEX86 proxy. Geochim. Cosmochim. Acta 127, 83–106 (2014).
- Frieling, J. et al. Extreme warmth and heat-stressed plankton in the tropics during the Paleocene-Eocene Thermal Maximum. Sci. Adv. 3, e1600891 (2017).
- Frieling, J. et al. Tropical Atlantic climate and ecosystem regime shifts during the Paleocene–Eocene Thermal Maximum. Clim. Past 14, 39–55 (2018).
- Bohaty, S. M., Zachos, J. C., Florindo, F. & Delaney, M. L. Coupled greenhouse warming and deep-sea acidification in the middle Eocene. *Paleoceanography* 24, PA2207 (2009).
- Sluijs, A., Pross, J. & Brinkhuis, H. From greenhouse to icehouse; organicwalled dinoflagellate cysts as paleoenvironmental indicators in the Paleogene. Earth Sci. Rev. 68, 281–315 (2005).
- Liu, Z. et al. Global cooling during the Eocene-Oligocene climate transition. Science 323, 1187–1190 (2009).
- Caballero, R. & Langen, P. L. The dynamic range of poleward energy transport in an atmospheric general circulation model. *Geophys. Res. Lett.* 32, L02705 (2005)



- Goldner, A., Herold, N. & Huber, M. Antarctic glaciation caused ocean circulation changes at the Eocene-Oligocene transition. *Nature* 511, 574–577 (2014); erratum 519, 378 (2015).
- Sluijs, A. et al. Warm and wet conditions in the Arctic region during Eocene Thermal Maximum 2. Nat. Geosci. 2, 777–780 (2009).
- Kiehl, J. T. & Shields, C. A. Sensitivity of the Palaeocene–Eocene Thermal Maximum climate to cloud properties. *Philos. Trans. R. Soc. A* 371, 20130093 (2013).
- Pierrehumbert, R. T. Thermostats, radiator fins, and the local runaway greenhouse. J. Atmos. Sci. 52, 1784–1806 (1995).
- Matthews, K. J. et al. Global plate boundary evolution and kinematics since the late Paleozoic. Global Planet. Change 146, 226–250 (2016).

Acknowledgements In this research, we used samples and data provided by the International Ocean Discovery Program (IODP) and its predecessor, the Ocean Drilling Program. This work was carried out under the programme of the Netherlands Earth System Science Centre (NESSC), financially supported by the Dutch Ministry of Education, Culture and Science. A.S. thanks the European Research Council (ERC) for ERC Starting Grant number 259627 under the European Union Seventh Framework Program. This study was made possible by the Netherlands Organisation for Scientific Research (NWO) grant number 334.11.006, which enabled the purchase of the UHPLC-MS system used for GDGT analyses. P.K.B. and F.P. acknowledge NWO-ALW Veni grants number 863.13.002 and number 863.13.016, respectively. M.H. was funded by the US National Science Foundation (NSF) grant OCE-0902882; the CESM model is also supported by the NSF. C.A. acknowledges the University of Padova grant number BIRD161002. We thank B. Wade (University College London) for converting the Eocene TDP data to GTS2012, P. Sexton (The Open University) for converting the middle Eocene Site 1258 data to GTS2012, J. Kiehl and

C. Shields for providing their simulation data, G.-J. Reichart (Royal NIOZ and Utrecht University) for discussions, L. van der Heijden (Utrecht University, now at University of La Rochelle, France), M. Nicolai (Utrecht University), A. Mets (NIOZ), N. Welters, A. van Dijk and D. Kasjaniuk (Utrecht Geolab) for analytical support.

Reviewer information *Nature* thanks G. Inglis, C. Lear, K. Littler and S. Robinson for their contribution to the peer review of this work.

Author contributions M.J.C., A.S. and M.H. designed the study. M.J.C., F.P., S.S., I.J.K., J.F. and E.L.K. generated and analysed organic geochemical data. M.J.C., A.S., P.K.B., J.F., I.J.K. and E.L.K. generated and analysed palynological data. C.A. generated and analysed nannofossil data. I.J.K., F.J.H., C.A., J.F., R.v.d.P. and M.J.C. developed the age model. M.H. and A.G. performed CESM model simulations. All authors contributed to data and model interpretations. M.J.C., A.S. and M.H. wrote the text, with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0272-2.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0272-2.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.J.C. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Palynology. Freeze-dried sediments (96 samples) were crushed and treated with 30% HCl and twice with 38%–40% HF to remove carbonates and silicates, respectively, after a known amount of *Lycopodium* spores (batch number 1031; 20,848 spores per tablet) was added to enable absolute quantification of palynomorphs. A 15–250 μm fraction was isolated using nylon mesh sieves and an ultrasonic bath. No oxidation procedure was applied. An aliquot of homogenized residue was mounted on slides and analysed using light microscopy (400 \times magnification) to a minimum of 200 identified dinocysts.

Organic geochemistry. Lipids were extracted from freeze-dried and powdered sediments (5-25 g dry weight, 118 samples) with dichloromethane (DCM):methanol (MeOH) (9:1, v/v) using a Dionex accelerated solvent extractor (ASE 350) at a temperature of 100 °C and a pressure of 7.6 \times 10 6 Pa. Lipid extracts were separated into an apolar, ketone and polar fraction by Al₂O₃ column chromatography using hexane:DCM (9:1), hexane:DCM (1:1) and DCM:MeOH (1:1) as respective eluents. 99 ng of a synthetic C_{46} (mass-to-charge ratio, m/z = 744) glycerol dialkyl glycerol tetraether (GDGT) standard was added to the polar fraction, which subsequently was dissolved in hexane:isopropanol (99:1, v/v) to a concentration of \sim 3 mg ml⁻¹ and passed through a 0.45- μ m polytetrafluoroethylene filter. This fraction was then analysed by high-performance liquid chromatography (HPLC) and atmospheric pressure chemical ionization-mass spectrometry using an Agilent 1260 Infinity series HPLC system coupled to an Agilent 6130 single-quadrupole mass spectrometer at Utrecht University following Hopmans et al.³¹ to measure the abundance of GDGTs. The branched and isoprenoid tetraether (BIT) index and $\rm TEX_{86}$ values were calculated according to Hopmans et al. 32 and Kim et al. 18, respectively. Based on long-term observation of the in-house standard, the analytical precision for TEX₈₆ is ± 0.3 °C.

 $\mbox{\bf GDGT}$ distributions. Of the 118 samples analysed for GDGTs, 5 early Eocene samples did not yield sufficient concentrations of GDGTs to determine TEX86. Additionally, 4 samples were excluded because either GDGT-2 (2 samples) or crenarchaeol (2 samples) could not be reliably identified. For the remaining 109 samples we evaluated the sources of GDGTs and the reliability of TEX86. The BIT index³², a means of quantifying the relative abundance of soil- and river-derived GDGTs relative to marine GDGTs, is low throughout the entire Eocene (all < 0.25, with 90% of values < 0.07) and there is no significant correlation between BIT index and TEX $_{86}$ (P > 0.3). Thus, our TEX $_{86}$ values are probably not biased by terrestrial input. Both the methane index and GDGT-2/Cren ratio show normal marine values (<0.20 and <0.12, respectively), so there is no indication for high abundance of methanotrophic archaea relative to Thaumarchaeota^{33,34}. Furthermore, GDGT-0/Cren is low (<0.31), so there are no indications for enhanced contributions of methanogenic archaea to the pool of isoprenoid GDGTs used in TEX₈₆³⁵ Finally, GDGT-2/GDGT-3 ratios are <4.5, ruling out substantial impact of deep-water production of GDGTs³⁶. Together, these ratios indicate that GDGT distributions were probably not considerably affected by either GDGT-producing soil bacteria, methanotrophic or methanogenic archaea, or deep-dwelling Thaumarchaeota, thereby designating upper-water-column Thaumarchaeota as the main source and favouring the interpretation of TEX₈₆^H as an SST proxy³⁷. Another recently described ratio focuses on the different GDGT distributions produced by modern Thaumarchaeota in the Red Sea⁵. Based on core-top datasets, fractional abundances of Red Sea GDGTs are known to differ from other oceanic settings, notably in containing relatively more crenarchaeol regio-isomer (Cren')³⁸. This causes a different relationship between TEX₈₆ and SST. Inglis et al.⁵ proposed the %GDGT_{RS}, [Cren'/(GDGT - 0 + Cren')] \times 100%, as a means of evaluating whether a 'Red Sea-type' GDGT distribution was present in the geological record. In our tropical Eocene record, TEX₈₆ is strongly driven by fractional abundance of Cren', therefore there is also a strong correlation between TEX $_{\!86}$ and %GDGT $_{\!RS}.$ However, as Inglis et al.⁵ noted, this Red Sea GDGT distribution cannot be distinguished from a high-temperature (>30 °C) distribution, so %GDGT_{RS} cannot disentangle the effects of high-temperature versus Red Sea-type GDGT distributions at this site. However, we note that the several reasons that have been proposed for the aberrant Red Sea GDGT distribution are not likely to have played a role at Site 959. There is no environmental similarity between Eocene Site 959 and the modern Red Sea that could account for a similarly adapted population of endemic Thaumarchaeota, as the setting is not comparable oceanographically or geomorphologically. Furthermore, our dinocyst record shows no indication of high salinity or strong stratification throughout the record. On this basis, we conclude that there is no reason to assume a similar relationship between TEX₈₆ and SST for Site 959 and the modern Red Sea. Finally, we note that our new equatorial record shows late Palaeocene TEX $_{86}^{H}$ SST estimates of 31–33 °C, identical to time-equivalent SSTs derived from the $\delta^{18}O$ and Mg/Ca ratios and TEX $_{86}^{H}$ of glassy preserved Morozovella acuta from nearby sections in Nigeria²⁰, confirming accurate proxy-estimated SSTs at the study site.

TEX₈₆ calibrations. Calibrations. Different calibrations have been proposed to translate TEX₈₆ into SST. Of note is also a recent paper by Ho and Laepple³⁹,

who propose that the sedimentary GDGTs derive from the deep ocean and TEX86 therefore reflects deep (>500 m) subsurface temperatures rather than SST. However, their conclusions are controversial, as their assumptions are inconsistent with all modern-ocean and microbiological evidence and the statistical method used is questionable ^40. Within the TEX $_{86}$ -to-SST calibrations, a first division can be made between calibrations based on core-top samples and those based on mesocosm experiments. Here, we focus on applying different calibrations based on core-top datasets^{18,41}, as these implicitly include ecological, water-column and diagenetic effects that are not incorporated in mesocosm experiments. Several linear and nonlinear core-top calibrations have been developed. Of these, the global nonlinear (logarithmic) TEX $_{86}^{\rm H}$ calibration of Kim et al. 18 and the BAYSPAR TEX $_{86}$ calibration of Tierney and Tingley^{19,41} are particularly applicable and most commonly chosen for higher-temperature settings, such as the Eocene. By treating TEX₈₆ as the dependent variable, BAYSPAR is the only calibration that does not suffer from regression dilution bias. For these calibrations, the differences in absolute temperature and relative temperature change in studies reporting TEX86 values between 0.5 and 0.75 are mostly within the error of the proxy 18,19,41 . Significant differences only appear with TEX_{86} values above those occurring in modern oceans (that is, $TEX_{86} > 0.73$) for which the TEX_{86} -to-SST calibration has to be extrapolated. This is illustrated in Extended Data Fig. 2a, which shows that SST estimates based on the TEX $_{86}^{H}$ and BAYSPAR calibration for Site 959 are within error between TEX₈₆ values of 0.67 and 0.80. However, the difference between the calibrations increases at higher TEX86 values. For assessing temperature change in a high temperature setting such as the equatorial Eocene, the choice of calibration therefore becomes an important factor.

Biophysical considerations. For the modern ocean, a linear calibration results in a better statistical correspondence between TEX₈₆ and SST in the temperature range of 5–30 $^{\circ}\text{C}^{18,42}.$ However, the question remains as to whether a linear calibration is the best choice for much warmer Eocene oceans considering the biochemical mechanism underlying the TEX86-SST relationship. Hyperthermophilic archaea in culture synthesize an increasing proportion of GDGTs with an increasing number of cyclopentane moieties with increasing temperature^{43–45}, probably as a homeoviscous adaptation of the cell membrane⁴⁶. However, the GDGTs included in the TEX₈₆ ratio (GDGT1-3 and the crenarchaeol isomer; see equation (1)) constitute a minor part of the membrane lipids of Thaumarchaeota. The dominant GDGTs are GDGT-0 and crenarchaeol⁴⁷⁻⁴⁹. Indeed, in the global core-top dataset, higher crenarchaeol and lower GDGT-0 are recorded with higher temperatures, although their response is less strong than that of the GDGT isomers included in TEX $_{86}$ (Fig. 4 in Kim et al. 18). Thus, TEX₈₆ does not capture the full membrane adaptation of Thaumarchaeota to changing temperatures. Interestingly, the ratio of crenarchaeol/ GDGT-0 versus TEX86 shows a strongly nonlinear relationship in the global coretop data (Extended Data Fig. 2c). This trend is similar to that observed between the TEX₈₆ and the ring index (RI; Extended Data Fig. 2d), which is the average number of cyclopentane rings of GDGTs 0-3, crenarchaeol and its regio-isomer (see equation (2)) and which also shows a strong relationship to temperature 50 .

$$TEX_{86} = \frac{[GDGT-2] + [GDGT-3] + [Cren']}{[GDGT-1] + [GDGT-2] + [GDGT-3] + [Cren']} \tag{1}$$

$$RI = 0 \times [GDGT-0] + 1 \times [GDGT-1] + 2 \times [GDGT-2]$$

$$+3 \times [GDGT-3] + 4 \times [Cren + Cren']$$
(2)

In Extended Data Fig. 2c, d the Eocene data from Site 959 overlap the core-top dataset, on both Red Sea and tropical latitude core-top data. This nonlinear relationship indicates that at high temperatures, TEX₈₆ shows a relatively small response to temperature change relative to the amount of crenarchaeol versus GDGT-0 and RI. This suggests that with increasing temperatures, adaptation of the thaumarchaeotal membrane is increasingly regulated through crenarchaeol and GDGT-0 rather than the GDGTs included in TEX_{86} . This should lead to a flattening of the slope between SST and TEX₈₆ and therefore supports a logarithmic relationship. Additional support for this hypothesis comes from recent culturing experiments on three different Thaumarchaeota strains⁴⁹. These show that for two strains, Nitrosopumilus maritimus and strain NAOA6, both TEX86 and RI (mainly driven by GDGT-0 and crenarchaeol) correlate with the incubation temperature. However, in the third strain (NAOA2), RI—but not TEX₈₆—changes with growth temperature. This third strain had the highest growth temperature optimum and the strongest change in RI from 28 °C to 35 °C. This suggests that at temperatures >28 °C, membrane adaptation to temperature in certain (hightemperature) Thaumarchaeota may not be well reflected in the TEX86 ratio. It should be noted that no nonlinear response was found in mesocosm experiments⁵¹ at temperatures of up to 40 °C. However, this calibration is substantially different from that of the global core-top calibrations owing to the unusually low amounts of the crenarchaeol regio-isomer⁵². A similarly low abundance of the crenarchaeol

regio-isomer was noted for *Nitrosopumilus maritimus* and strain NAOA6⁴⁹. In strain NAOA2⁴⁹, abundances of crenarchaeol regio-isomer were higher and did increase with temperature, suggesting that it may be a better representation for high-temperature-adapted marine Thaumarchaeota. On the basis of the above biophysical evidence, we argue that the slope of the TEX_{86} -to-temperature curve is likely to flatten at temperatures above the surface sediment dataset, such as the TEX_{86} values recorded in the Eocene of Site 959.

Implications. The use of the nonlinear TEX_{86}^H calibration results in lower temperature estimates compared to the linear BAYSPAR calibration (Extended Data Fig. 2b) for Site 959 in the early Eocene and late Palaeocene. Notably, for the Palaeocene, such estimates agree better with SSTs derived from glassy preserved planktonic foraminiferal δ^{18} O and Mg/Ca records from nearby sections in Nigeria²⁰. Additionally, the Site 959 TEX_{86}^{H} estimates fit well with the other SST constraints that we use in our tropical Eocene compilation (Extended Data Fig. 5). Finally, the similarly reduced sensitivity of TEX₈₆ (that is, nonlinearity) at the low end of the temperature range is undisputed because it is apparent in both mesocosm experiments⁵² and in the global core-top dataset^{18,53}. We therefore apply the TEX₈₆ calibration in our main analysis, which is presented in the main text. Nevertheless, the absolute temperature estimates and magnitude of change obtained from the extrapolated part of the TEX₈₆ calibration curve should always be interpreted with care. For completeness, we also present the results for the MTG and polar amplification analysis using BAYSPAR in Extended Data Fig. 3. This confirms that the use of TEX $^{\rm H}_{86}$ instead of BAYSPAR gives a conservative estimate of middle late Eocene cooling and MECO warming at Site 959 and thus a low estimate of (early Eocene) MTGs and a maximum estimate of polar amplification compared to BAYSPAR. Crucially, however, the choice of calibration does not affect the trends in tropical surface temperatures (Extended Data Fig. 3a) or the fact that they parallel deep-ocean temperatures, and therefore does not affect our conclusion regarding the drivers of Eocene climate change. The larger Eocene range of tropical temperatures reconstructed using BAYSPAR does imply that SSTs at Site 959 varied more than deep-sea temperatures during the Eocene, suggesting tropical rather than polar amplification. Regardless, the relation between tropical and deepsea temperatures remains linear, reflecting a constant polar amplification factor (Extended Data Fig. 3c).

Age model Site 959. Eocene sediments from Site 959 were too weakly magnetized to yield reliable palaeomagnetic results¹⁶. Our age model is therefore based on a combination of bio- and chemostratigraphy and supported by cyclic variations in sediment coloration (Extended Data Fig. 1). Although dinoflagellate cyst assemblages support the Eocene age of the analysed material, they do not yield many biostratigraphic events with a well calibrated age in the tropics⁵⁴. A total number of 76 additional standard smear slides (Supplementary Information) were analysed for calcareous nannofossils and enabled the improvement of the initial biostratigraphic framework⁵⁵. Biochronological estimates from the low-latitude nannofossil biozonation⁵⁶ were converted to GTS2012⁵⁷ using the relative position of each biohorizon within the respective magnetochron. In total, 10 robust nannofossil tie-points were used (Extended Data Table 1). The base and top of Chiasmolithus gigas could not be used at this site owing to the extremely low abundance of this species. Therefore, alternative biohorizons in the evolutionary lineage Sphenolithus furcatolithoides morph. A-Sphenolithus cuniculus-S. furcatolithoides morph. B were used. On the basis of the co-occurrence of two non-synchronous bio-events (base Nannotetrina alata gr. and base Nannotetrina cristata) at the same depth (between 740.95 and 741.63 metres below sea floor (mbsf)) and supported by a sudden shift in the nannofossil assemblage, the presence of a hiatus was inferred in Core 35R at \sim 741 mbsf. The presence of *Nannotetrina alata sensu strico* in combination with Sphenolithus perpendicularis and transitional forms of sphenoliths at 740.95 mbsf suggests that the sediments just above the hiatus are very close in age to the actual base of the N. alata group. Therefore, we also include this biohorizon in our age model. The lower boundary of the hiatus is based on linear extrapolation of the underlying sedimentation rate of 1.27 cm kyr⁻¹. This approach results in a hiatus of 1.5 Myr (48.0-46.5 Myr ago). To further constrain the age model, several chemostratigraphic tie-points were used. The onset of the carbon isotope excursion marking the Palaeocene–Eocene thermal maximum (~56 Myr ago) was recently identified²¹ at 804.09 mbsf. In addition, the previously identified late Eocene minimum in osmium isotope ratios⁵⁸ (¹⁸⁷Os/¹⁸⁸Os) at 458.65 mbsf has an age of 34.4 Myr ago in GTS2012 on the basis of the correlation to the Os isotope record at the well dated ODP sites 1218 and 1219⁵⁹. These age constraints indicate that our data span the entire Eocene. The age model is further supported by calculated sedimentation rates from selected intervals, where high-resolution colour logs showed more than four easily distinguishable cycles. Sedimentation rates were calculated by assuming that these smallest-scale alternations are precession-forced, and were thus assigned a duration of 21 kyr per cycle. These sedimentation rates (blue lines in Extended Data Fig. 1) correspond closely to those based on chemostratigraphy and biostratigraphy. Our age model implies that the warming interval 590-565 mbsf reflects the MECO, which is further supported by a shift in osmium isotope ratios that was also identified within the MECO at sites 1263 and U1333⁶⁰. Owing to a lack of nannofossils in the poorly recovered upper part of Hole 959D, linear extrapolation was used for the data points below 466 mbsf. This places the Eocene–Oligocene boundary at 447.5 mbsf, which is in good agreement with the placing of the Oi1 glacial event on the basis of osmium isotope recovery after the minimum of 34.4 Myr ago by Ravizza and Paquay⁵⁸.

Age models other sites. *ODP Site 1172*. The TEX $_{86}$ -based SST record from Site $1172^{3,61,62}$ is plotted (Fig. 2) on an age model based on the magnetostratigraphy of Bijl et al. ⁶³, which is in turn largely grounded on the original interpretation by Fuller and Touchard ⁶⁴. This was supplemented with three well-calibrated dinocyst events from Dallanave et al. ⁶⁵ (top and base *Charlesdowniea edwardsii* and top *Wilsonidium ornatum*) instead of the uncertain magnetochron reversals for this interval (552–578 mbsf).

Dahomey Basin, Nigeria. For the Sagamu Quarry and IB10B Core, Nigeria, published biostratigraphic and chemostratigraphic age constraints²⁰ were used. Specifically, base Morozovella subbotinae and base Acarinina soldadoensis were used for the Sagamu Quarry (SQ) and base Acarinina soldadoensis, top Morozovella acuta, carbon isotope excursion (CIE) onset and top CIE recovery were used for IB10B as age–depth tie-points.

Tropical SST compilation. For the presented compilation, we integrate the new ODP Site 959 TEX₈₆-based SST record with several existing SST proxy records, specifically δ¹⁸O of photosymbiont-bearing planktonic foraminifera *Morozovella* spp. (upper mixed layer) and Acarinina spp. (mixed layer) from the SQ, Nigeria²⁰ and TDP sections^{4,66} and near-surface dwelling *Turborotalia ampliapertura* from TDP⁶⁷; Mg/Ca of Morozovella spp. from ODP Site 865⁶⁸ and SQ²⁰, Acarinina spp. from SQ²⁰ and T. ampliapertura from TDP⁶⁹; TEX₈₆ from ODP Site 925^{24,70}, Site 929^{5,24}, TDP⁴, SQ and the IB10B Core, Nigeria²⁰; and clumped isotope (Δ_{47}) thermometry of shallow-dwelling large benthic foraminifera from Evans et al.⁶ (Supplementary Information). We did not include data from South Dover Bridge⁵ and Walvis Ridge⁷¹ because plate tectonic reconstructions place them outside the $30^{\circ}\,N\text{--}30^{\circ}\,S$ latitude band. Age models for all sites were converted to GTS2012 using published age-depth tie-points. For the Mg/Ca proxy, (normalized) Mg/Ca compositions were converted to SST using the calibration from Anand et al.⁷² and the Eocene seawater Mg/Ca reconstruction from Evans et al.⁶, and using⁷³ H = 0.42 to correct for the power-law dependence of test Mg/Ca values on changing seawater Mg/Ca ratios⁷⁴. Conversion of δ^{18} O to temperature was done following Erez and Luz⁷⁵, assuming a constant ice-free global $\delta^{18}O_{sw}$ of -1.2% VPDB⁷⁶ (Vienna Pee Dee belemnite) and (constant) latitudinal corrections for TDP and SQ of 0.83% and 0.61%, respectively⁷⁷. A +2 °C correction to convert reconstructed T. ampliapertura temperatures to SST (as used in the original publication⁶⁹) was omitted here. It should be noted that different seawater chemistry assumptions for the δ^{18} O and Mg/Ca proxies may result in shifts in reconstructed temperatures, but do not qualitatively change trends or the correspondence between trends. Multiple measured specimens per sample in the original studies have been averaged into one value for this compilation. For the $\mathrm{TEX}_{86}\text{-based}$ records, samples with aberrant GDGT ratios were removed following Inglis et al.⁵. The logarithmic TEX H calibration of Kim et al. 18 is presented in the main text and a full supplementary analysis using the linear BAYSPAR¹⁹ calibration is provided. For the BAYSPAR calibration, the default search tolerance (2 standard deviations of the raw TEX₈₆ dataset) was used for Site 959, 925 and 929, which yields a representative set of low-latitude calibration localities. The search tolerance was stretched to 0.15 TEX units for the Dahomey Basin and TDP records, in order to not only sample the possibly anomalous modern Red Sea³⁸, but also include a broad representative sample of low-latitude localities.

Global deep-ocean temperature compilation. We base our deep-ocean temperature compilation on the benthic isotope stack previously compiled by Zachos et al. 2 and supplement this with several high-resolution benthic isotope records, specifically from ODP Site 690 78,79 , ODP Site 748 22 , ODP Site 1218 80 , ODP Site 1209 81 , ODP Site 1258 82 , ODP Site 1262 83 and ODP Site 1263 84 . After the respective species-specific corrections for disequilibrium vital effects 85 were applied, δ^{18} O-to-temperature conversion was done following Erez and Luz 75 , assuming a constant ice-free global δ^{18} Osw of -1.2% VPDB 76 . Age models for all sites were converted to GTS2012 using published age–depth tie-points.

 ${
m CO_2}$ compilation. The compiled ${
m CO_2}$ record plotted in Fig. 2 derives from boron isotopes from TDP^{8,86}, with 68% confidence intervals, as reported in Foster et al. ⁸⁷, and alkenones from ODP Site 612⁸⁸ and Site 925⁷⁰.

Meridional temperature gradients. As an approximation for the pole-to-Equator temperature difference or MTG, we calculate the difference between tropical mean SST and deep-ocean temperatures. The latter are better constrained than high-latitude SSTs and exclude potential summer temperature biases that might plague available high-latitude SST records. We use deep-ocean temperatures based on δ^{18} O, as these are better constrained than those based on Mg/Ca, particularly because of the large uncertainties regarding seawater Mg/Ca values as well as larger uncertainty between different calibrations and corrections used to convert

Mg/Ca to temperature. We note that our approach of using the LOESS-fitted data provides robust estimates of long-term changes in MTG, but is less appropriate for considering transient events, as fitted event MTGs (for example, for the PETM and MECO) are very dependent on the bandwidth of the fit and the specific records used (Extended Data Fig. 8).

Sensitivity to late Eocene ice volume. The Cenozoic benthic foraminiferal δ^{18} O signal reflects both deep-water temperature and global ice volume changes. Although it is unlikely that large Antarctic ice sheets were present in the warmest interval of the Cenozoic during the early Eocene, the extent of middle late Eocene Antarctic glaciation is more uncertain (see, for example, Miller et al. 89, Barker et al. 90 and Gasson et al. 91). Recent work argues for possible early middle Eocene glaciation⁹² but the dating of these sediments is highly uncertain. Although there might be evidence for glacial activity, the interpreted presence of large East Antarctic ice sheets in the early middle Eocene is highly controversial, certainly in light of very warm temperatures on the East Antarctic margin⁹³. Nevertheless, initial small ice caps in the middle late Eocene would have had relatively enriched isotopic compositions of -20% to -35% VSMOW⁹⁴ (Vienna Standard Mean Ocean Water) relative to mean modern Antarctic ice (-54% VSMOW). We assess the effect of a range of middle late Eocene ice volumes with different isotopic compositions on the mean δ^{18} O of Eocene seawater (Extended Data Fig. 6d). This demonstrates that the effect of ice volume was probably not more than \sim 0.25‰, or $\sim\!\!1\,^{\circ}\text{C}$, in the latest Eocene. To further illustrate this, we present both the record of ice-free deep-ocean temperature evolution and a second line based on a linear build-up of late Eocene ice volume from 39.5 Myr ago (post-MECO) onwards to a latest-Eocene (34.0 Myr ago) maximum of 10⁷ km³ (refs^{95,96}) with an isotopic composition of -25% VSMOW⁹⁴ (Extended Data Fig. 6a). This makes a maximum difference of about 0.8 °C ($\Delta\delta^{18}O_{sw}$ of 0.18%) in the latest Eocene. We further propagate this uncertainty into the analysis of MTGs and polar amplification factors (Extended Data Fig. 6b, c).

CESM 1 model simulations. The CESM 1 simulations share the same generalized Eocene palaeogeography to assess the effect of changing CO₂ by itself, and were all run for more than 3,000 yr to equilibrium. Simulations using an earlier, and generally similar, version of this model were found to produce the best match to early Eocene proxy temperatures within a multi-model ensemble¹⁵, and preliminary comparison revealed that these new simulations are slightly improved over the earlier version for the early Eocene²⁰. Results from the lower-CO₂ simulations (560 p.p.m. and 1,120 p.p.m.) and further information on the model can be found in Goldner et al. 26. This version of CESM has a modern 'fast' climate sensitivity of 2.9 °C per CO₂ doubling⁹⁷ and a nearly constant 'slow' climate sensitivity (ESS) of 3.5 °C per CO₂ doubling in the Eocene simulations used here. For comparison with the proxies in this study, the four simulations with varying CO2 were assigned specific ages by matching the simulated deep-ocean temperatures to the proxy-based deep-ocean temperature reconstruction curve. We then compared the resulting SSTs at the same localities as the main sites in our proxy compilation (ODP sites 865, 925/929 and 959 and TDP) and surface-to-deep gradients to evaluate model performance (Supplementary Information). The temperature at the proxy data localities was sampled in a 4° radius. This approach avoids the circularity of adjusting the climate model radiative forcing to match surface temperature records and provides a target that circumvents the uncertainty introduced by the various errors and uncertainties in surface temperature records. In these simulations, bottom water temperatures in the 4,480 p.p.m. scenario (simulation EO4) are representative of a hot early Eocene climatic optimum extreme (deep-ocean temperature of 13-14 °C, following the conventions of Huber and Caballero 14) whereas the 560 p.p.m. scenario (simulation EO1) is comparable to the latest Eocene (deepocean temperature of 4-5 °C), with intermediate simulations (EO2 and EO3) being in between and comparable to the middle Eocene.

Polar amplification factor calculations. First, to obtain an estimate of the factor by which polar temperature change is amplified relative to the tropics (that is, the polar amplification factor), we performed a Deming regression of the Site 959 record against the deep-sea stack of temperatures, accounting for errors in both variables. Data were binned into 1-Myr bins from 34 to 58 Myr ago. We did not include data from the EOT and earliest Oligocene, to exclude major effects of ice volume changes on seawater $\delta^{18}O$. To assess the robustness of the single regression, we followed a probabilistic approach, using Monte Carlo resampling with full propagation of errors. First, we generated 1,000 iterations of both the tropical SST and deep-ocean temperature datasets. In these iterations, each data point was resampled within the 95% confidence limits of its propagated analytical plus calibration uncertainty, assuming Gaussian distribution of errors. Using these, we performed 1,000 iterations of a Deming regression of deep-ocean temperature against tropical SST, with data binned into 1-Myr bins from 34 to 58 Myr ago and propagated errors related to the binning used in the regression. We plotted the resulting suite of 1,000 slopes as a probability density function of the polar amplification factor. This exercise was performed using the full tropical temperature compilation and the single Site 959 record. We additionally performed a Deming regression of

the Site 959 record against the high-latitude Site 1172 record as a supplementary analysis. The latter analysis generates a similar polar amplification factor, but with larger scatter and uncertainty. This is due to the smaller amount of data points in the Site 1172 SST record relative to the benthic δ^{18} O stack, as well as differences in the detailed pattern of Eocene cooling between Site 1172 compared to Site 959 and the deep-ocean temperature record.

ESS calculations. To provide estimates of Eocene ESS sensu Lunt et al. 98, we combined our proxy and model reconstructions of temperature with the few available CO₂ reconstructions based on boron isotopes⁸, involving 1 sample for the early (54-49 Myr ago), 2 samples for the middle (48-42 Myr ago) and 1 sample for the late (38-35 Myr ago) Eocene. We derived temperatures by sampling the proxy compilation within the designated age brackets. We use tropical and deep-ocean temperature change (dT) as minimum and maximum estimates of dT. Between these, a uniform 'flat' probability distribution was assumed. We converted changes in boron-based CO₂ estimates to radiative forcing in W m⁻² using the radiative forcing fit from Byrne and Goldblatt⁹⁹. With the above approach, we derived estimates of ESS in K W⁻¹ m⁻² for the early Eocene compared to the middle and late Eocene and to the preindustrial temperature. Uncertainties are based on propagated uncertainties of temperature change and radiative forcing derived by resampling these datasets 1,000 times within their 95% confidence limits (propagated analytical plus calibration uncertainty for temperature, reported 95% confidence limits from the original work⁸ for CO₂). In this, we removed radiative forcings <0, that is, we assumed that there is no negative forcing associated with increasing CO_2 . Given the good match between proxies and the presented model simulations, we also calculated ESS using the model-derived global mean temperatures and CO₂ proxy data.

Data availability. The data supporting the findings of this study are available within the paper and its Supplementary Information. Original raw data (palynology counts and GDGT concentrations and chromatograms) are available from the corresponding author upon request.

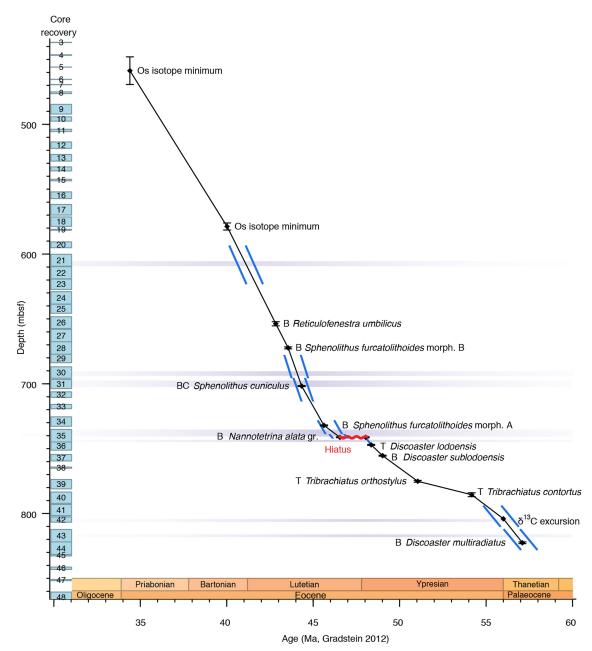
Code availability. The model used in this study is NCAR CESM 1 with CAM4 atmosphere, which is freely available from NCAR (http://www.cesm.ucar.edu/models/cesm1.0/).

- 31. Hopmans, E. C., Schouten, S. & Sinninghe Damsté, J. S. The effect of improved chromatography on GDGT-based palaeoproxies. *Org. Geochem.* **93**, 1–6 (2016).
- Hopmans, E. C. et al. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. Earth Planet. Sci. Lett. 224, 107–116 (2004)
- Weijers, J. W. H., Lim, K. L. H., Aquilina, A., Sinninghe Damsté, J. S. & Pancost, R. D. Biogeochemical controls on glycerol dialkyl glycerol tetraether lipid distributions in sediments characterized by diffusive methane flux. Geochem. Geophys. Geosystems 12, Q10010 (2011).
- Zhang, Y. G. et al. Methane index: a tetraether archaeal lipid biomarker indicator for detecting the instability of marine gas hydrates. *Earth Planet. Sci. Lett.* 307, 525–534 (2011).
- Blaga, C. I., Reichart, G.-J., Heiri, O. & Damsté, J. S. S. Tetraether membrane lipid distributions in water-column particulate matter and sediments: a study of 47 European lakes along a north–south transect. J. Paleolimnol. 41, 523–540 (2009).
- Taylor, K. W. R., Huber, M., Hollis, C. J., Hernandez-Sanchez, M. T. & Pancost, R. D. Re-evaluating modern and Palaeogene GDGT distributions: implications for SST reconstructions. Global Planet. Change 108, 158–174 (2013).
- Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S.
 Distributional variations in marine crenarchaeotal membrane lipids: a new tool
 for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* 204,
 265–274 (2002).
- Trommer, G. et al. Distribution of Crenarchaeota tetraether membrane lipids in surface sediments from the Red Sea. Org. Geochem. 40, 724–731 (2009).
- Ho, S. L. & Laepple, T. Flat meridional temperature gradient in the early Eocene in the subsurface rather than surface ocean. *Nat. Geosci.* 9, 606–610 (2016).
- 40. Tierney, J. E., Sinninghe Damsté, J. S., Pancost, R. D., Sluijs, A. & Zachos, J. C. Eocene temperature gradients. *Nat. Geosci.* **10**, 538 (2017).
- Tierney, J. E. & Tingley, M. P. A. TEX86 surface sediment database and extended Bayesian calibration. Sci. Data 2, 150029 (2015).
- Kim, J.-H., Schouten, S., Hopmans, E. C., Donner, B. & Sinninghe Damsté, J. S. Global sediment core-top calibration of the TEX86 paleothermometer in the ocean. Geochim. Cosmochim. Acta 72, 1154–1173 (2008).
- De Rosa, M., Gambacorta, A., Nicolaus, B. & Bu'Lock, J. D. Complex lipids of Caldariella acidophila, a thermoacidophile archaebacterium. *Phytochemistry* 19, 821–825 (1980).
- Lai, D., Springstead, J. R. & Monbouquette, H. G. Effect of growth temperature on ether lipid biochemistry in Archaeoglobus fulgidus. *Extremophiles* 12, 271–278 (2008).
- Boyd, E. S. et al. Temperature and pH controls on glycerol dibiphytanyl glycerol tetraether lipid composition in the hyperthermophilic crenarchaeon Acidilobus sulfurireducens. Extremophiles 15, 59–65 (2011).
- Gliozzi, A., Paoli, G., De Rosa, M. & Gambacorta, Á. Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaebacteria. *Biochim. Biophys. Acta* 735, 234–242 (1983).



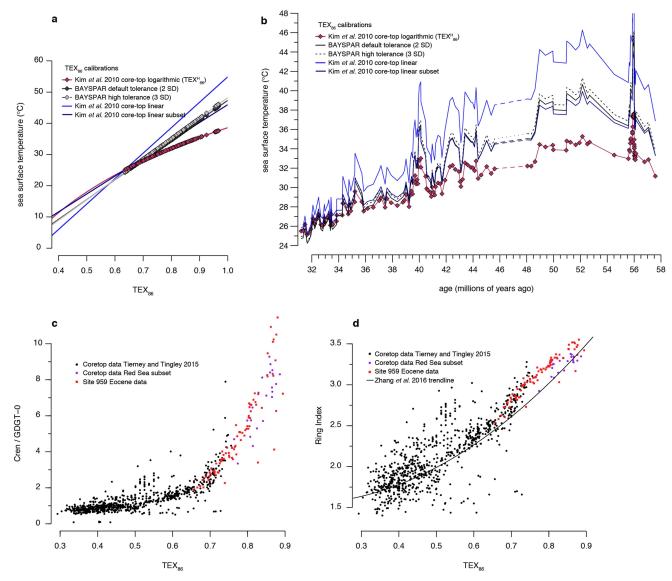
- Schouten, S. et al. Intact membrane lipids of "Candidatus Nitrosopumilus maritimus," a cultivated representative of the cosmopolitan mesophilic group I crenarchaeota. *Appl. Environ. Microbiol.* 74, 2433–2440 (2008).
- Pitcher, A. et al. Core and intact polar glycerol dibiphytanyl glycerol tetraether lipids of ammonia-oxidizing archaea enriched from marine and estuarine sediments. Appl. Environ. Microbiol. 77, 3468–3477 (2011).
- Elling, F. J., Könneke, M., Mußmann, M., Greve, A. & Hinrichs, K.-U. Influence of temperature, pH, and salinity on membrane lipid composition and TEX86 of marine planktonic thaumarchaeal isolates. *Geochim. Cosmochim. Acta* 171, 238–255 (2015).
- Zhang, Y. G., Pagani, M. & Wang, Z. Ring Index: a new strategy to evaluate the integrity of TEX86 paleothermometry. *Paleoceanography* 31, 220–232 (2016).
- Schouten, S., Forster, A., Panoto, F. E. & Sinninghe Damsté, J. S. Towards calibration of the TEX86 palaeothermometer for tropical sea surface temperatures in ancient greenhouse worlds. *Org. Geochem.* 38, 1537–1546 (2007).
- Wuchter, C., Schouten, S., Coolen, M. J. L. & Sinninghe Damsté, J. S.
 Temperature-dependent variation in the distribution of tetraether membrane
 lipids of marine Crenarchaeota: implications for TEX86 paleothermometry.
 Paleoceanography 19, PA4028 (2004).
- Ho, S. L. et al. Appraisal of TEX86 and thermometries in subpolar and polar regions. Geochim. Cosmochim. Acta 131, 213–226 (2014).
- Awad, W. K. & Oboh-Ikuenobe, F. E. Early Paleogene dinoflagellate cysts from ODP Hole 959D, Côte d'Ivoire-Ghana Transform Margin, West Africa: new species, biostratigraphy and paleoenvironmental implications. *J. Afr. Earth Sci.* 123, 123–144 (2016).
- Shafik, S., Watkins, D. K. & Shin, I. C. Calcareous nannofossil paleogene biostratigraphy, Côte d'Ivoire-Ghana Marginal Ridge, Eastern Equatorial Atlantic. Proc. Ocean Drill. Program Sci. Results 159, 413–431 (1998).
- Agnini, C. et al. Biozonation and biochronology of Paleogene calcareous nannofossils from low and middle latitudes. Newsl. Stratigr. 47, 131–181 (2014).
- Gradstein, F. M., Ogg, J. G., Schmitz, M. D. & Ogg, G. M. The Geologic Time Scale 2012 (Elsevier, Amsterdam, 2012).
- Ravizza, G. & Paquay, F. Os isotope chemostratigraphy applied to organic-rich marine sediments from the Eocene-Oligocene transition on the West African margin (ODP Site 959). *Paleoceanography* 23, PA2204 (2008).
- Dalai, T. K., Ravizza, G. E. & Peucker-Ehrenbrink, B. The late Eocene ¹⁸⁷Os/¹⁸⁸Os excursion: chemostratigraphy, cosmic dust flux and the Early Oligocene glaciation. *Earth Planet. Sci. Lett.* **241**, 477–492 (2006).
- van der Ploeg, R. et al. Middle Eocene greenhouse warming facilitated by diminished weathering feedback. Nat. Commun. (in the press).
- Bijl, P. K. et al. Transient middle Eocene atmospheric CO₂ and temperature variations. Science 330, 819–821 (2010).
- Sluijs, A. et al. Southern ocean warming, sea level and hydrological change during the Paleocene-Eocene thermal maximum. Clim. Past 7, 47–61 (2011).
- Bijl, P. K., Sluijs, A. & Brinkhuis, H. A magneto- and chemostratigraphically calibrated dinoflagellate cyst zonation of the early Palaeogene South Pacific Ocean. *Earth Sci. Rev.* 124, 1–31 (2013); erratum 134, 160–163 (2014).
 Fuller, M. & Touchard, Y. in *The Cenozoic Southern Ocean: Tectonics*,
- Fuller, M. & Touchard, Y. in The Cenozoic Southern Ocean: Tectonics, Sedimentation, and Climate Change Between Australia and Antarctica (eds Exon, N. F. et al.) 63–78 (American Geophysical Union, Washington DC, 2004).
- Dallanave, E. et al. Constraining early to middle Eocene climate evolution of the southwest Pacific and Southern Ocean. Earth Planet. Sci. Lett. 433, 380–392 (2016).
- Áze, T. et al. Extreme warming of tropical waters during the Paleocene–Eocene Thermal Maximum. Geology 42, 739–742 (2014).
- Pearson, P. N. et al. Extinction and environmental change across the Eocene-Oligocene boundary in Tanzania. Geology 36, 179–182 (2008).
- Tripati, A. K. et al. Tropical sea-surface temperature reconstruction for the early Paleogene using Mg/Ca ratios of planktonic foraminifera. *Paleoceanography* 18, 1101 (2003).
- Lear, C. H., Bailey, T. R., Pearson, P. N., Coxall, H. K. & Rosenthal, Y. Cooling and ice growth across the Eocene-Oligocene transition. *Geology* 36, 251–254 (2008).
- Zhang, Y. G., Pagani, M., Liu, Z., Bohaty, S. M. & DeConto, R. A. 40-million-year history of atmospheric CO₂. Philos. Trans. R. Soc. A 371, 20130096 (2013).
- Boscolo Galazzo, F. et al. The middle Eocene climatic optimum (MECO): a multiproxy record of paleoceanographic changes in the southeast Atlantic (ODP Site 1263, Walvis Ridge). Paleoceanography 29, 1143–1161 (2014).
- Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* 18, 1050 (2003).

- Hasiuk, F. J. & Lohmann, K. C. Application of calcite Mg partitioning functions to the reconstruction of paleocean Mg/Ca. Geochim. Cosmochim. Acta 74, 6751–6763 (2010).
- Evans, D. & Müller, W. Deep time foraminifera Mg/Ca paleothermometry: nonlinear correction for secular change in seawater Mg/Ca. *Paleoceanography* 27, PA4205 (2012).
- Erez, J. & Luz, B. Experimental paleotemperature equation for planktonic foraminifera. Geochim. Cosmochim. Acta 47, 1025–1031 (1983).
- Shackleton, N. J. & Kennett, J. P. Paleotemperature history of the Cenozoic and the initiation of Antarctic glaciation: oxygen and carbon isotope analyses in DSDP Sites 277, 279 and 281. *Initial Rep. Deep Sea Drill. Proj.* 29, 743–755 (1975)
- Žachoś, J. C., Stott, L. D. & Lohmann, K. C. Evolution of Early Cenozoic marine temperatures. *Paleoceanography* 9, 353–387 (1994).
- Kennett, J. P. & Stott, L. D. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. *Nature* 353, 225–229 (1991).
- Thomas, D. J., Záchos, J. C., Bralower, T. J., Thomas, E. & Bohaty, S. Warming the fuel for the fire: evidence for the thermal dissociation of methane hydrate during the Paleocene–Eocene thermal maximum. *Geology* 30, 1067–1070 (2002).
- Coxall, H. K. & Wilson, P. A. Early Oligocene glaciation and productivity in the eastern equatorial Pacific: insights into global carbon cycling. *Paleoceanography* 26, PA2221 (2011).
- Westerhold, T., Röhl, U., Donner, B., McCarren, H. K. & Zachos, J. C. A complete high-resolution Paleocene benthic stable isotope record for the central Pacific (ODP Site 1209). *Paleoceanography* 26, PA2216 (2011).
- Sexton, P. F. et al. Eocene global warming events driven by ventilation of oceanic dissolved organic carbon. *Nature* 471, 349–352 (2011).
- Littler, K., Röhl, U., Westerhold, T. & Zachos, J. C. A high-resolution benthic stable-isotope record for the South Atlantic: implications for orbital-scale changes in Late Paleocene–Early Eocene climate and carbon cycling. Earth Planet. Sci. Lett. 401, 18–30 (2014).
- Lauretano, V., Hilgen, F. J., Zachos, J. C. & Lourens, L. J. Astronomically tuned age model for the early Eocene carbon isotope events: a new high-resolution δ¹³C benthic record of ODP Site 1263 between ~49 and ~54 Ma. Newsl. Stratigr. 49, 383–400 (2016).
- Shackleton, N. J. & Hall, M. A. The late Miocene stable isotope record, Site 926. Proc. Ocean Drill. Program, Sci. Results 154, 367–373 (1997).
- Pearson, P. N., Foster, G. L. & Wade, B. S. Atmospheric carbon dioxide through the Eocene–Oligocene climate transition. *Nature* 461, 1110–1113 (2009).
- Foster, G. L., Royer, D. L. & Lunt, D. J. Future climate forcing potentially without precedent in the last 420 million years. *Nat. Commun.* 8, 14845 (2017).
- Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. Science 309, 600–603 (2005).
- Miller, K. G., Wright, J. D. & Browning, J. V. Visions of ice sheets in a greenhouse world. Mar. Geol. 217, 215–231 (2005).
- Barker, P. F., Diekmann, B. & Escutia, C. Onset of Cenozoic Antarctic glaciation. Deep Sea Res. Part II Top. Stud. Oceanogr. 54, 2293–2307 (2007).
- Gasson, E. et al. Exploring uncertainties in the relationship between temperature, ice volume, and sea level over the past 50 million years. Rev. Geophys. 50, RG1005 (2012).
- Gulick, S. P. S. et al. Initiation and long-term instability of the East Antarctic Ice Sheet. Nature 552, 225–229 (2017).
- Pross, J. et al. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. Nature 488, 73–77 (2012).
- DeConto, R. M. et al. Thresholds for Cenozoic bipolar glaciation. Nature 455, 652–656 (2008).
- Gasson, E. et al. Uncertainties in the modelled CO₂ threshold for Antarctic glaciation. Clim. Past 10, 451–466 (2014).
- De Boer, B., van de Wal, R. S. W., Bintanja, R., Lourens, L. J. & Tuenter, E. Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic 8¹⁸O records. *Ann. Glaciol.* 51, 23–33 (2010).
- 97. Shields, C. A. et al. The low-resolution CCSM4. J. Clim. 25, 3993–4014 (2012)
- Lunt, D. J. et al. Earth system sensitivity inferred from Pliocene modelling and data. Nat. Geosci. 3, 60–64 (2010).
- Byrne, B. & Goldblatt, C. Radiative forcing at high concentrations of well-mixed greenhouse gases. Geophys. Res. Lett. 41, 152–160 (2014).
- Torsvik, T. H. et al. Phanerozoic polar wander, palaeogeography and dynamics. Earth Sci. Rev. 114, 325–368 (2012).



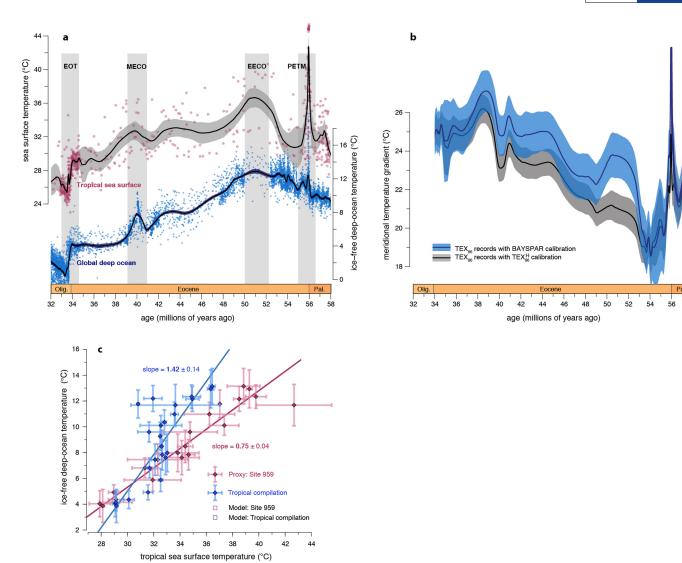
Extended Data Fig. 1 | **Augmented age model of Hole 959D.** Age-depth plot showing calcareous nannofossil and chemostratigraphic tie-points (diamonds; vertical error bars are indicate the minimum and maximum depth of the tie-point), as presented in Extended Data Table 1b. B, BC and

T stand for base, base common and top, respectively. Blue-shaded regions represent depth intervals for which sedimentation rates (blue lines) were calculated. The hiatus of $\sim\!1.5$ Myr in Core 35 is indicated as a red curly line. Epochs and ages are expressed in Myr ago (Ma), following GTS2012.



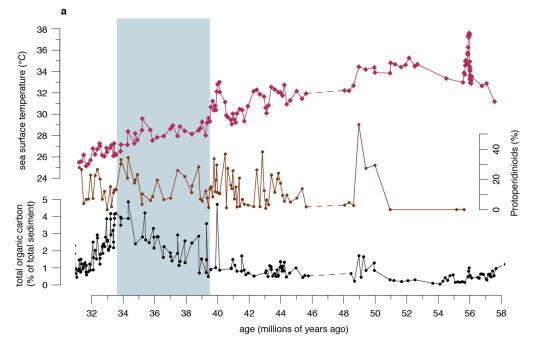
Extended Data Fig. 2 | Comparison between different TEX₈₆-to-SST calibrations and different GDGT ratios. a, TEX₈₆-SST calibration lines (trend lines for BAYSPAR) for one logarithmic and several linear calibrations. Plotted symbols are the Site 959 TEX₈₆ record, to illustrate which part of the calibration is relevant for this study. Compared calibrations are: BAYSPAR^{19,41} with default settings (search tolerance for 2 TEX₈₆ standard deviations, 0.13; dark-grey line, dark-grey diamonds), BAYSPAR with increased search tolerance (0.2) (dashed line, light-grey diamonds), Kim et al. ¹⁸ logarithmic TEX₈₆ core-top calibration (red line, red diamonds), linear core-top calibration ¹⁸ (light-blue line) and linear subset core-top calibration without Red Sea and polar ocean data ¹⁸ (dark-

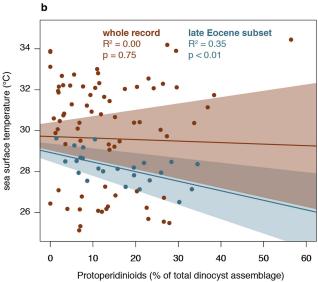
blue line). It is of note that the logarithmic TEX $_{86}^{\rm H}$ starts strongly diverging from the linear BAYSPAR and subset calibrations from TEX $_{86}$ values of >0.8. **b**, Site 959 SST record using different TEX $_{86}$ calibrations. Calibrations and line colours and types are as in **a**. **c**, Ratio of crenarchaeol to GDGT-0 against TEX $_{86}$. Data are from a core-top compilation⁴¹ (black circles; Red Sea subset, purple circles) and our Site 959 record (red squares). **d**, Ring index sensu Zhang et al. ⁵⁰ against TEX $_{86}$. Data are from a core-top compilation (black circles; Red Sea subset, purple circles) and our Site 959 record (red squares). The exponential regression line of Zhang et al. through the core-top data is plotted as a black line.

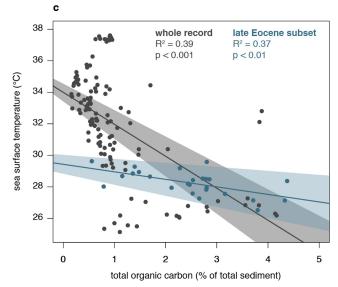


Extended Data Fig. 3 | Sensitivity of main results to TEX₈₆ calibration. a, Top, tropical SST compilation. Proxy data are compiled as described in Methods (red symbols), with all TEX₈₆-based records converted to SST using the BAYSPAR calibration (default settings, search tolerance as described in Methods). The fitted LOESS model is plotted as a black line, with the 95% confidence interval as grey shading. Bottom, ice-free deepocean temperature compilation. $\delta^{18}\text{O-based}$ proxy data are compiled as described in Methods. The fitted LOESS model is plotted as a black line and the 95% confidence interval as dark blue shading. b, Calculated MTG based on LOESS fits of proxy data (lines, propagated 95% confidence intervals shown as shading). The black line with the grey silhouette shows the tropical compilation with TEX $_{86}^{H}$ calibration and the blue line with the

blue silhouette shows the BAYSPAR calibration. c, Proxy (blue diamonds, tropical compilation; red diamonds, Site 959) deep-ocean temperature against tropical SST using the BAYSPAR calibration for TEX86-based records. Lines represent Deming regression analysis through proxy data. The slope (polar amplification factor) is $1.42\pm0.14~(\pm1~{\rm standard~error})$ for the tropical compilation and 0.75 ± 0.04 for Site 959. Proxy data grouped into 1-Myr bins from 34–58 Myr ago, with error bars representing one standard deviation due to binning. This sensitivity analysis shows that calculated MTGs and the constant polar amplification factor are relatively robust to the specific TEX86 calibration used, although MTGs are less reduced in the early Eocene when using the BAYSPAR calibration. Polar amplification factors are lower, but reflect a linear relationship.

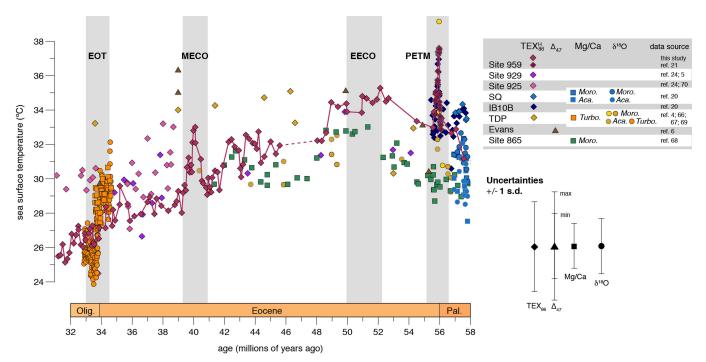






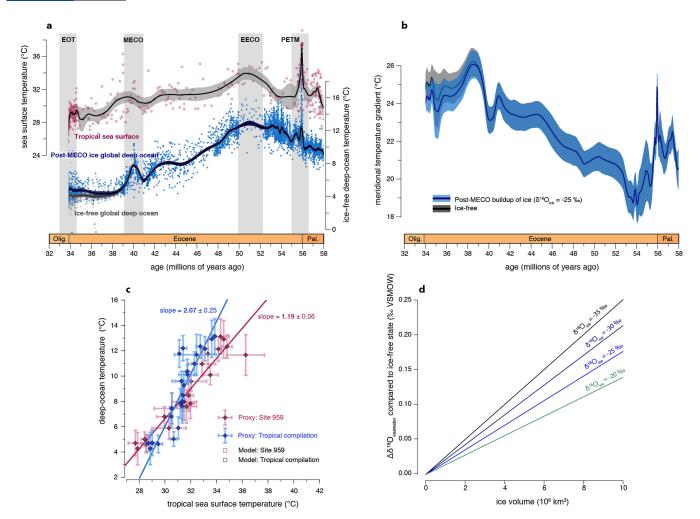
Extended Data Fig. 4 | Regression analysis between reconstructed SST and abundance of upwelling indicators. a, $\rm TEX^{\rm H}_{86}$ -based SST (red diamonds, upper left vertical axis), protoperidinioid abundance (percentage of total dinocyst assemblage; brown dots, right vertical axis) and TOC (percentage of sediment; black dots, lower left vertical axis) records of ODP Site 959. Dashed lines represent a hiatus in Site 959. Age is in GTS2012. b, Regression analysis between SST and percentage of protoperidinioid dinocysts of total dinocyst assemblage, showing a nonsignificant relationship with a very low fit (brown line, 90% confidence

interval shown as brown shading; $R^2=0.00$, P=0.75) and a better fit ($R^2=0.35$) that is significant (P<0.01) when only the late Eocene (post-MECO) part of the record is considered (blue–grey line; 90% confidence interval is shown as blue–grey shading). **c**, Regression analysis between SST and percentage of TOC in sediment, showing a significant negative correlation for the whole record ($R^2=0.39$, P<0.001; dark-grey line, with the 90% confidence interval shown as dark-grey shading) and the late Eocene subset ($R^2=0.37$, P<0.01; blue–grey line, with the 90% confidence interval shown as blue–grey shading).



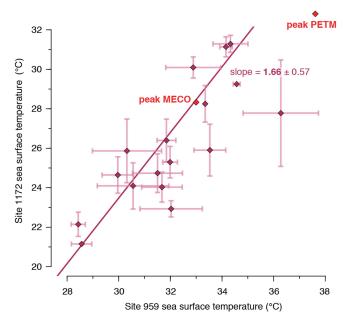
Extended Data Fig. 5 | **Eocene multi-proxy tropical SST compilation.** Compilation presented in Fig. 2, here plotted per site and proxy, with data sources in the key. The abbreviations *Moro.*, *Aca.*, and *Turbo*. stand for foraminifera genera *Morozovella*, *Acarinina* and *Turborotalia*, respectively. The dashed line in the Site 959 record represents a hiatus. Conservative estimates of propagated calibration and analytical errors (1 s.d.) are

 $\pm 2.6\,^{\circ}\text{C}$ for TEX $_{86}^{H},\pm 1.6\,^{\circ}\text{C}$ for $\delta^{18}\text{O}$ and $\pm 1.3\,^{\circ}\text{C}$ for Mg/Ca. Δ_{47} uncertainties are as reported in the original study 6 , with the minimum and maximum per-sample uncertainty. Uncertainties are plotted on the same relative vertical temperature scale as the data to facilitate comparison. The age is in Myr ago, following GTS2012.

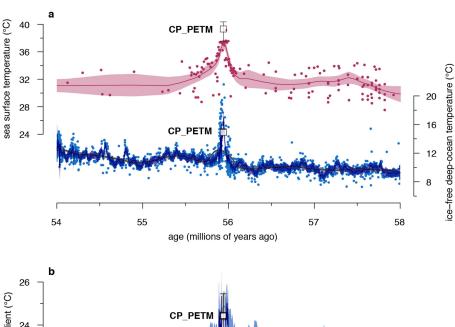


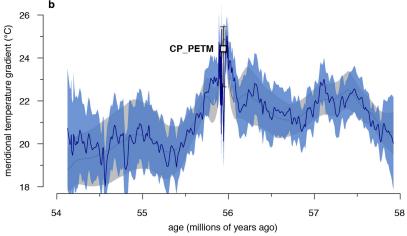
Extended Data Fig. 6 | Sensitivity of main results to late Eocene ice volume. a, Top, tropical SST compilation; proxy data are compiled as described in Methods (red symbols). The fitted LOESS model is plotted as a black line and the 95% confidence interval as grey shading. Bottom, deep-ocean temperature compilation; δ^{18} O-based proxy data are compiled as described in Methods. Ice-free deep-ocean temperatures and fitted LOESS model are shown as grey dots and line, respectively, and the deep-ocean temperature compilation and fitted LOESS model including late Eocene ice volume effect (Methods) as blue dots and line, correspondingly. 95% LOESS confidence intervals are shown as shading. b, Calculated MTG based on LOESS fits of proxy data (lines; propagated 95% confidence intervals are shown as silhouettes). The black line with grey silhouette

shows results obtained using ice-free deep-ocean temperatures, and the blue line with blue silhouette includes the late Eocene ice volume effect on the deep-ocean temperature. **c**, Proxy (blue diamonds, tropical compilation; red diamonds, Site 959) deep-ocean temperature, including the late Eocene ice volume effect, against tropical SST. Lines represent Deming regression analysis through proxy data. The slope (polar amplification factor) is $2.07\pm0.25~(\pm1~\text{standard error})$ for the tropical compilation and 1.19 ± 0.06 for Site 959). Proxy data grouped into 1-Myr bins from 34–58 Myr ago, with error bars representing one standard deviation due to binning. **d**, Sensitivity of δ^{18} O of Eocene seawater (‰ VSMOW) to the build-up of 0–10 $^7~\text{km}^3$ of ice with varying isotopic composition.



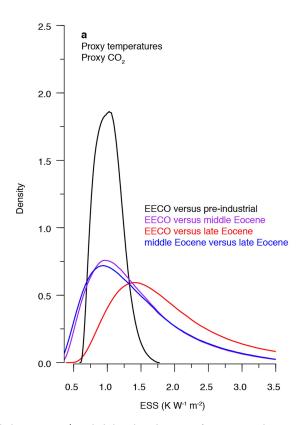
Extended Data Fig. 7 | Linear relationship between high-latitude and tropical SST. Site 1172 TEX $_{86}$ -based SST (record plotted in Fig. 2) against Site 959 TEX $_{86}$ -based SST. Lines represent Deming regression analysis through proxy data (polar amplification factor, 1.66 ± 0.57). Proxy data are grouped into 1-Myr bins from 34–58 Myr ago, with error bars representing one standard deviation due to binning. Peak PETM and peak MECO SSTs are plotted as separate points, which fall within the uncertainty of the regression line.

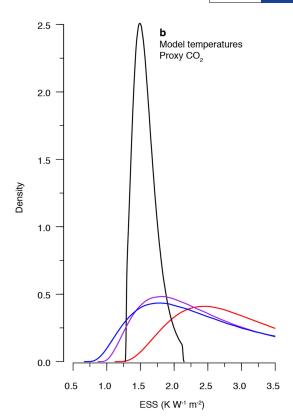




Extended Data Fig. 8 | PETM temperature gradient proxy-model comparison. a, Top, tropical SST compilation; proxy data are compiled as described in Methods (red symbols). The fitted LOESS model is plotted as a red line, with the 95% confidence interval as pink shading. Bottom, ice-free deep-ocean temperature compilation; δ^{18} O-based proxy data are compiled as described in Methods. Generalized cross-validation (GCV)-optimized fitted LOESS model (as in Fig. 3) plotted as grey line, with the 95% confidence interval as grey shading. An alternative LOESS model with small bandwidth (0.25 times the GCV-optimized span) that tracks deepocean PETM temperature more closely is shown as the blue line, with the 95% confidence interval as blue shading. Data are plotted together with

the PETM simulation of Kiehl and Shields²⁸ (black open squares, seasonal range error bars) with altered cloud parameters (CP_PETM). **b**, Calculated MTG based on LOESS fits of proxy data and of the model simulation CP_PETM. The grey line is obtained using a GCV-optimized fitted LOESS model and the blue line using a smaller-bandwidth deep-ocean LOESS model (propagated 95% confidence intervals are shown as shading). We note that the PETM MTG pattern is complex owing to its sensitivity to the specific records, age models and filtering used, and might have evolved over the course of the event. Nevertheless, peak PETM MTG matches the simulation CP_PETM poorly. The age is in Myr ago, following GTS2012.





Extended Data Fig. 9 | Probability distributions of Eocene Earth system sensitivity. a, b, ESS estimates using proxy (a) and model (b) temperatures in combination with proxy-based $\rm CO_2$ concentrations, derived as described in Methods. Eocene ESS is separated into the late Eocene relative to the EECO (red), the middle Eocene relative to the EECO (purple) and

the middle Eocene relative to late Eocene (blue). ESS estimates of the EECO relative to preindustrial temperature (black) have lower error owing to the high precision of preindustrial $\rm CO_2$ concentration and temperature, but include additional long-term non- $\rm CO_2$ effects.



Extended Data Table 1 \mid Palaeolatitude and age constraints of Site 959 over the Eocene

а

Palaeolatitude at	Matthews et al. 2016	Torsvik <i>et al.</i> 2012		
60 Ma	0.982 °N	9.826 °S		
55 Ma	0.779 °N	9.480 °S		
50 Ma	0.581 °N	9.130 °S		
45 Ma	0.587 °N	7.405 °S		
40 Ma	0.600 °N	5.674 °S		
35 Ma	0.942 °N	4.568 °S		
30 Ma	1.291 °N	3.462 °S		

b

Event	Species / Proxy	Minimum depth (mbsf)	Maximum depth (mbsf)	Mean depth (mbsf)	One- sided error (m)	Age GTS2012	Reference
Latest Eocene Os isotope minimum	¹⁸⁷ Os/ ¹⁸⁸ Os	448.00	469.30	458.65	10.65	34.40	Ravizza and Paquay 2008(⁵⁸)
MECO Os isotope minimum	¹⁸⁷ Os/ ¹⁸⁸ Os	576.05	581.51	578.78	2.73	40.02	van der Ploeg <i>et al.</i> in press(⁶⁰)
Base	Reticulofenestra umbilicus	652.12	655.34	653.73	1.61	42.84	Shafik <i>et al.</i> 1998(⁵⁵); this study
Base	Sphenolithus furcatolithoides morph. B	671.56	673.10	672.33	0.77	43.55	This study
Base common	Sphenolithus cuniculus	701.38	702.12	701.75	0.37	44.33	Shafik <i>et al</i> . 1998; this study
Base	Sphenolithus furcatolithoides morph. A	731.73	732.49	732.11	0.38	45.63	This study
Base	Nannotetrina alata group	740.95	741.63	741.29	0.34	46.55	Shafik <i>et al.</i> 1998; this study
Bot hiatus	-	740.95	741.63	741.29	0.34	48.03	This study
Тор	Discoaster Iodoensis	746.73	747.45	747.09	0.36	48.36	This study
Base	Discoaster sublodoensis	755.06	755.80	755.43	0.37	49.01	Shafik <i>et al</i> . 1998; this study
Тор	Tribrachiatus orthostylus	774.49	775.86	775.18	0.69	51.05	This study
Тор	Tribrachiatus contortus	784.05	787.03	785.54	1.49	54.17	Shafik <i>et al.</i> 1998
Onset PETM isotope excursion	δ ¹³ C	804.08	804.10	804.09	0.01	56.00	Frieling <i>et al.</i> 2018(²¹)
Base	Discoaster multiradiatus	821.92	823.26	822.59	0.67	57.10	Shafik <i>et al.</i> 1998

a, Palaeolatitudes reconstructed with GPlates using the hotspot reference frame of Matthews et al.³⁰ and the palaeomagnetic reference of Torsvik et al.¹⁰⁰. Present latitude is 3.6276° N and longitude is 2.7352° W. **b**, Bio- and chemostratigraphic age-depth tiepoints (from refs ^{21,55,58,60} and this work) used in developing the age model for the Eocene of Site 959.



Global surface warming enhanced by weak Atlantic overturning circulation

Xianyao Chen¹ & Ka-Kit Tung²*

Evidence from palaeoclimatology suggests that abrupt Northern Hemisphere cold events are linked to weakening of the Atlantic Meridional Overturning Circulation (AMOC)¹, potentially by excess inputs of fresh water². But these insights—often derived from model runs under preindustrial conditions—may not apply to the modern era with our rapid emissions of greenhouse gases. If they do, then a weakened AMOC, as in 1975–1998, should have led to Northern Hemisphere cooling. Here we show that, instead, the AMOC minimum was a period of rapid surface warming. More generally, in the presence of greenhouse-gas heating, the AMOC's dominant role changed from transporting surface heat northwards, warming Europe and North America, to storing heat in the deeper Atlantic, buffering surface warming for the planet as a whole. During an accelerating phase from the mid-1990s to the early 2000s, the AMOC stored about half of excess heat globally, contributing to the global-warming slowdown. By contrast, since mooring observations began³⁻⁵ in 2004, the AMOC and oceanic heat uptake have weakened. Our results, based on several independent indices, show that AMOC changes since the 1940s are best explained by multidecadal variability6, rather than an anthropogenically forced trend. Leading indicators in the subpolar North Atlantic today suggest that the current AMOC decline is ending. We expect a prolonged AMOC minimum, probably lasting about two decades. If prior patterns hold, the resulting low levels of oceanic heat uptake will manifest as a period of rapid global surface warming.

As an analogy of the flow of energy in our climate system, consider the filling of a bucket of water from a tap at the top. The feed rate of the tap is an analogue of the top-of-atmosphere radiative imbalance—the net heating—of our planet, with the water level in the bucket analogous to surface warming. The sink at the bucket bottom drains into a larger bucket below (the deeper oceans). If the drain rate is the same as the feed rate from the tap at the top, the water level in the bucket does not rise (hiatus of surface warming). If the drain is plugged, the water level will rise rapidly in the bucket (rapid surface warming). AMOC controls about half of the variation of this 'drain rate'.

Figure 1 quantifies the energy budget of our climate system, using the subsurface ocean heat content (OHC) measured mostly by a system of autonomous profiling Argo floats, during a period, 2000-2014, when the 'drain rate' was large. The total OHC, as approximated by that in the upper 1,500 m of the oceans, is increasing at a rate of about 0.42 ± 0.02 W m $^{-2}$, consistent with radiative imbalance 7 . The upper 200 m roughly corresponds to the mixed layer globally. Through wind and turbulent mixing, variations of sea surface temperature (SST) and mixed-layer OHC are highly statistically correlated (r = 0.82 in 13-month running mean). Figure 1 shows that both were in a warming slowdown for this period. Why the upper 200 m OHC was in a warming slowdown is clear: the increase in heat storage below 200 m, about 89 zettajoules $(1 \text{ ZJ} = 10^{21} \text{ J})$. This amount of heat is equivalent to 180 years of the world's energy consumption at the current rate, and any future variation even within this observed range will have important consequences for the surface temperature.

If the radiative imbalance and the heat storage below 200 m were to remain the same, the 0–1,500 m OHC would still increase at the same rate as the radiative imbalance, but the 0–200 m OHC curve would lie on the 0–1,500 m curve, increasing at the same rate, or about 0.23 °C per decade. Our best estimate for the next two decades, allowing for some increase in ocean storage, is 70% of that rate, at 0.16 °C per decade (see Methods), close to the 25-year trend of 0.177 °C per decade of the last rapid warming period in the twentieth century.

The inset of Fig. 1 shows how the global increase in OHC storage between 200 m and 1,500 m are partitioned among the various oceans. The Pacific and the Indian oceans dominate the horizontal exchanges of heat in the upper 300 m^{9,10}, and the Atlantic and the Southern oceans dominate the vertical redistribution¹¹. They accounted for about 70% of the global heat storage increase in the 200–1,500 m layer during 2000–2014, divided between the North Atlantic, which is dominant before 2005, and the Southern Ocean after 2005. The subsurface warming in the Southern Ocean started in 1993 according to the data available (see below), and was attributed to the southward displacement and intensification of the circumpolar jet⁸, caused in large part by the Antarctic ozone hole¹². The North Atlantic's role appears to be cyclic on decadal timescales, with AMOC in an accelerating phase before 2005.

AMOC transports warm saline surface water found in the subtropical Atlantic to the subpolar Atlantic, where heat loss to the cold atmosphere increases its density. Aided by its high salinity it sinks and returns southward at depth. When AMOC is stronger (weaker), more (less) of the warm and saline water is found in the subpolar Atlantic, and subsequent sinking subducts more (less) heat there, as demonstrated in Fig. 2. The contrast is dramatic between periods when AMOC is increasing and when it is decreasing. Why AMOC sometimes accelerates or declines is more complicated. It could be responding to external forcing, for example, such as the freshening of the subpolar waters from melting ice at the end of the Little Ice Age¹³. Or, AMOC could be part of a natural, multidecadal variability involving feedbacks between the density effect of salinity on deep convection in Labrador and the Nordic Seas, and the subsequent induced northward transport of surface salinity reinforcing the deep convection¹⁴.

AMOC is commonly believed to be slowing on centennial timescales owing to global warming. The RAPID/MOCHA mooring array, deployed in 2004³ off the coast of Florida to monitor AMOC, soon afterwards recorded its weakening⁴. The decadal decline, however, is ten times larger than the predicted forced response⁵, causing concerns about its long-term trend and possible deficiencies of the models used. Figure 3a, constructed from various independent proxies from 1945 to the present (see Extended Data Fig. 1 for unfiltered time series and Extended Data Fig. 2 for error bars), shows that it is dominated instead by reversing phases. The weakening AMOC, by 3.7 Sverdrups (Sv) since 2005 measured by the RAPID/MOCHA array, was actually preceded by an acceleration¹5,16. Altimetry data of sea-surface heights (SSH) available since 1993¹¹ were used to deduce¹8 via geostrophic balance that at 4¹° N AMOC sped up by 4 Sv from the early 1990s to 2005, consistent with Zhang's subsurface fingerprint proxy⁶. We use multiple

¹Physical Oceanography Laboratory, CIMST, Ocean University of China, and Qingdao National Laboratory of Marine Science and Technology, Qingdao, China. ²Department of Applied Mathematics, University of Washington, Seattle, WA, USA. *e-mail: ktung@uw.edu

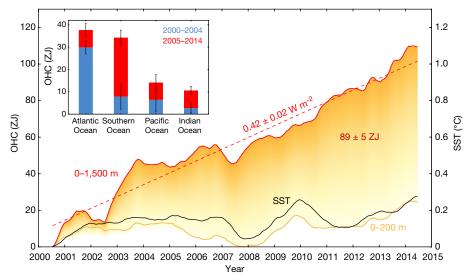


Fig. 1 | Quantifying the global heat budget and the partition among ocean basins in the two periods 2000–2004 and 2005–2014. The SST from ERSST.v4 is shown as a black curve and the 0–200-m OHC from the ISHII and Scripps datasets (see Methods) is shown as an orange curve, showing that they co-vary and that both are in a warming slowdown, while the total OHC, as approximated by the 0–1,500-m OHC (red curve), is increasing at the regressed linear rate of 0.42 W m $^{-2}$ (red dashed straight line). This excess heat from forcing is sequestered below 200 m. The orange-shaded region represents the additional amount of heat stored in the 200–1,500 m layer since 2000, about 89 ZJ. One zettajoule

is equivalent to twice the world's annual energy consumption. If this additional storage were absent, the upper 200 m would have increased at the rapid rate of the red curve. We adjusted the data for the Southern Ocean to remove a possible artefact due to the rapid transition from no-Argo to the Argo observing platform around $2002-2003^{28}$. The inset shows the division of the 89 ZJ of global ocean increase in heat storage in the 200-1,500 m layer into the four ocean basins and two periods. 35° S marks the northern boundary of the Southern Ocean and the southern boundary of the Atlantic, Pacific and Indian oceans. The error bars are one-standard-deviation errors of the linear regression.

independent proxies to infer subpolar AMOC strength back in time to 1945. Many of the proxies used here have been validated by models: Zhang's subsurface temperature fingerprint was highly coherent with AMOC strength^{6,19,20} at low frequencies in the model (GFDL CM2.1) at mid-latitudes. The subpolar gyre SST proxy²¹, and the upper ocean subpolar salinity proxy²⁰ were also model-validated. Along with the long record of tide gauges along the east coast of the USA²², these proxies

consistently indicate a period of low AMOC from the mid-1970s to the 1990s. The shading in Fig. 3 shows that this period coincided with a period of rapid surface warming. See also Extended Data Fig. 3 for the coincidence of Atlantic OHC change and global surface warming. See Methods for model–observation reconciliation.

We call AMOC+ (AMOC-) the phase when the AMOC strength is above (below) climatology (based on the subpolar salinity, which has

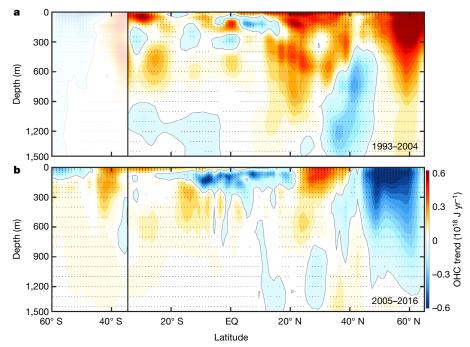


Fig. 2 | **The OHC linear trend in the Atlantic basin.** The trend is zonally averaged over two periods, when AMOC is increasing (**a**) and decreasing (**b**). The two periods are chosen according to the observed AMOC trends in Fig. 3a. ISHII data are used in the first period and Scripps data are used

in the second period. Stippling indicates areas of statistical significance at the 95% confidence level. The linear trend is unreliable in the Southern Ocean prior to 2005, and so that region is masked.

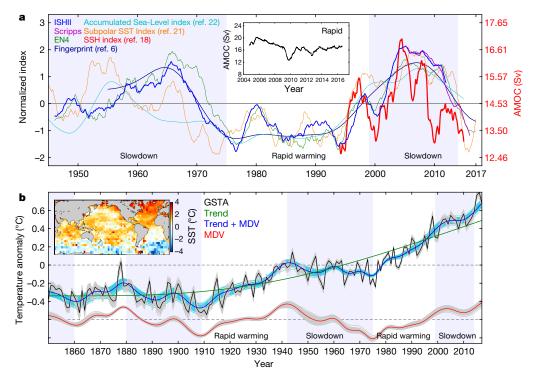


Fig. 3 | AMOC and GSTA variations. a, Mid and subpolar latitude AMOC strength, as calculated at 41° N using altimetry measurements, from ref. ¹⁸ (red, two-year running mean, Sverdrup scale shown on the right); inferred from integrated subpolar salinity in 0–1,500 m and 45–65° N in the Atlantic as a proxy, using the ISHII (dark blue) and Scripps (purple) datasets, with a two-year running mean. The green curve is the subpolar salinity, similarly calculated but using EN4. The AMOC fingerprint⁶ (dark blue) and the accumulated sea-level index (turquoise) calculated from historical tide gauge measurements²² were smoothed with 10-year and 7-year low-pass filters, respectively, from their sources. The subpolar gyre

SST index²¹ in orange is also a two-year running mean. See Methods for details. The inset shows RAPID-measured AMOC at 26° N. b, Shown are GSTA from HadCRUT4.6 (black), the nonlinear secular trend (close to the 100-year linear trend) (brown) and variation about the trend for timescales longer than decadal (multidecadal variability (MDV), red). The inset shows the SST spatial pattern associated with MDV obtained by regressing SST onto its time series. The blue curve is the smoothed version of GSTA obtained as the sum of the secular trend and MDV. The faint lines around the solid lines are from 100 ensemble members of the HadCRUT4.6, which assess the range of uncertainty of the data used in the solid lines.

a long record with no trend). The high (+) phase consists of two rapid subphases. The increasing subphase (AMOC_{up}) started in 1993, from the low point in AMOC-, first slowly and then rapidly, peaking in 2005. It is then followed by a rapid decreasing subphase (AMOC_{down}) (2005 to the present) (Fig. 3a). At low values of overturning (AMOC–) the strength is relatively level even though there are short-term fluctuations, because a slower poleward transport of saline water from the tropical Atlantic makes it difficult to speed up the sinking in the subpolar North Atlantic except through slower processes: The surface water could slowly become more saline through the reduction of fresh water outflow from land glaciers and from the Arctic Ocean²³. The northward transport of warm and saline water increased more rapidly since 1999, and started a negative feedback as the warm surface water increased glacier melt and freshwater outflow. The previous AMOC_{down} subphase of 1965–1974 started with the gradual freshening of the north Atlantic waters, as can be inferred from the decreasing salinity in the subpolar region, braking the AMOC. Incidentally, both SSH at 41° N and RAPID at 26° N showed a simultaneous, short-lived 30% drop in AMOC strength in 2009–2010⁵, partially caused by an extreme negative episode of atmospheric North Atlantic Oscillation that affected the wind field⁵ over both areas.

Water masses in the subpolar and subtropical gyres are different and transports across gyre boundaries need not be continuous ¹⁴. For vertical heat subduction, it is mainly the subpolar AMOC that is our focus in Fig. 3a. Signals from salinity proxies at the subpolar Atlantic have almost reached the previous low. The subpolar gyre SST has started to warm. The deep Labrador Sea density, which is known to lead by 7–10 years changes in wider basin AMOC^{15,16}, has stopped declining since 2014 (Extended Data Fig. 4). The subtropical region is more prone to higher-frequency perturbations ¹⁴, and the RAPID time series is

experiencing its short-term oscillations (two so far) after the recovery from the large dip in 2010 so the decadal trend may be difficult to see. Nevertheless, it appears to have stabilized at that latitude. Previously, when AMOC reached its lowest AMOC – value after 1975, that level phase lasted two and a half decades. Although we have data only for one cycle, its observed non-sinusoidal pattern characterized by a prolonged flat minimum separated by steep peaks is as expected from the physical arguments presented above.

The longer Global-mean Surface Temperature Anomaly (GSTA) record shown in Fig. 3b, together with its low-frequency variation^{24,25}, consists of a secular trend and a multidecadal variability (MDV), defined to be on timescales that are decadal or longer. The spatial pattern associated with MDV (inset to Fig. 3b) has the pattern of an interhemispheric seesaw in the Atlantic, with the North Atlantic being the centre of action, consistent with model results²⁶. When the MDV is increasing it doubles the GSTA warming rate over the 100-year trend of 0.08 K per decade, and is associated with a period of rapid warming in the late and also the early twentieth century. That secular trend of 0.08 K per decade, statistically significant at over 95% confidence level against a second-order autoregressive (AR(2)) red noise, has been attributed to the underlying anthropogenic global warming trend²⁷. The regressed spatial pattern associated with the secular trend resembles the model-predicted response from greenhouse warming^{24,25}. The MDV in the GSTA is related to the Atlantic Multidecadal Oscillation (AMO) (see Methods), the latter having a record extending back several hundred years.

The previous period of low overturning in the AMOC— phase, from 1975 to the 1990s, coincided with a period of rapid global warming at the surface. This is more than a coincidence because the energy budget involved can be quantified. We do not have reliable subsurface data for

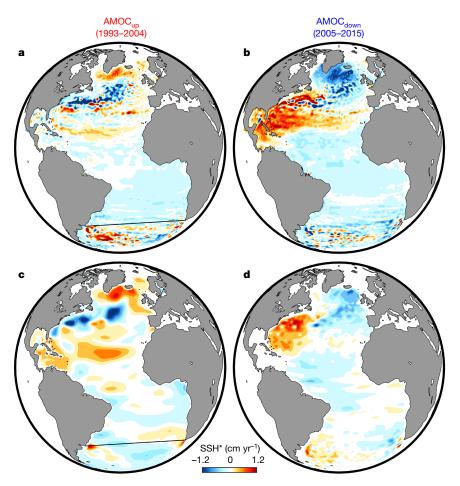


Fig. 4 | Contrasting thermosteric SSH* patterns for increasing and decreasing AMOC. a, c, Linear SSH* trend when AMOC is increasing; b, d, Linear SSH* trend when AMOC is decreasing. a and b show SSH*

from remote sensing, compared with the steric sea level calculated using OHC in **c** and **d**. SSH* is SSH with its global mean subtracted, reflecting mostly the thermosteric part of SSH (see Methods).

the period when the surface warming was rapid. However, the change from that period can be quantified so that an estimate can be made for what would happen if that change were absent. During 2000–2005, in the AMOC $_{\rm up}$ subphase, 52% of the global increase between 200 m and 1,500 m is sequestered in the Atlantic. Together with the heat sequestrated in the Southern Ocean, it contributed to a period of global warming slowdown. When this additional heat storage is absent, a period of rapid surface warming is expected to reoccur.

Although the Argo programme was launched around 2000, its coverage in the Southern Ocean did not become adequate until 2005. To validate the data on OHC we compare satellite SSH* (the asterisk indicates the deviation of SSH from its global mean) available since 1993 (Fig. 4a and b) to the thermosteric sea level rise (due to thermal expansion of the water column) (Fig. 4c and d) calculated using OHC above 1,500 m. The comparison is surprisingly good north of 35° S. Notable exceptions are as expected; they include areas with no Argo measurements: shallow maritime areas west of the Caribbean islands, and the deep mid-Atlantic Ocean below 1,500 m, which was not included in our OHC. South of 35° S the linear trend in the Argo data is not reliable across 2003 during the transition from no-Argo to Argo measurements²⁸. The two datasets consistently show that in the subpolar Atlantic there is increasing (decreasing) heat storage when AMOC is increasing (decreasing). The southward (northward) displacement of the Gulf Stream at midlatitudes created some compensating cooling (warming)²¹. In the AMOC's rapidly decreasing subphase, some heat is entrained in the subtropical gyre. The Southern Hemisphere north of 35° S is mostly featureless. South of 35° S, mesoscale patterns of warming can be seen in SSH*, which is also reflected in the OHC after 2004, but not before, owing to data quality. These mesoscale eddies in the linear trend occurring south of the Antarctic Circumpolar Current may be due to its recent strengthening, and its increased baroclinic instability²⁹.

The increased sea level (Fig. 4b) and warmer SST (Extended Data Fig. 5d) in the western subtropical Atlantic may have led to strong hurricanes and their destructive power, and the surprising string of category-5 hurricanes making landfall towards the end of the decreasing phase of the AMOC, instead of at the peak of the AMOC, when the mean SST of the entire North Atlantic is the warmest and the basin-wide hurricane number is the highest³⁰.

Climate-model runs under preindustrial conditions demonstrated the existence of multidecadal variation in AMOC, and its associated Atlantic SST variation: the AMOC+ (AMOC-) phase corresponds to warm (cold) SST and Northern Hemisphere mean surface temperature^{6,19}. This prevailing paradigm has permeated popular perceptions about the future climate consequence of an AMOC weakened by global warming, similar to the abrupt switch back into icy conditions of the Younger Dryas during the last deglaciation². Over the past few decades, however, there is a positive trend of warmer subsurface water in the subpolar Atlantic (Extended Data Fig. 6), rendering the mean state lighter (see the temperature-salinity diagram in Extended Data Fig. 7). Deep convections can now carry more heat downward. In the presence of greenhouse heating from above and warmer SSTs, AMOC's role in sequestering heat becomes important in the current global surface energy budget (Fig. 1). When AMOC is more constant, as in the AMOC – phase, little additional heat is sequestered in the Atlantic, contributing to a more rapid surface warming as more heat from radiative imbalance remains on the surface and the upper 200 m of the global oceans. We note, however, that we have discussed here only one component of a complex system: global heat balance is maintained by



the combined ocean and atmosphere systems and a change in the transport of one regional component may affect the partitioning of change between other parts of the ocean or of the atmosphere, depending on the timescales involved.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0320-y.

Received: 25 January 2018; Accepted: 6 June 2018; Published online 18 July 2018.

- McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. Nature 428, 834–837 (2004).
- Manabe, S. & Stouffer, R. J. Coupled ocean-atmosphere model response to freshwater input: comparison to Younger Dryas Event. Paleoceanography 12, 321-336 (1997).
- McCarthy, G. D. et al. Measuring the Atlantic Meridional Overturning Circulation at 26° N. Prog. Oceanogr. 130, 91-111 (2015).
- Smeed, D. A. et al. Observed decline of the Atlantic meridional overturning circulation 2004-2012. Ocean Sci. 10, 29-38 (2014).
- Srokosz, M. A. & Bryden, H. L. Observing the Atlantic Meridional Overturning Circulation yields a decade of inevitable surprises. Science 348, (2015).
- Zhang, R. Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. Geophys. Res. Lett. 35, L20705 (2008).
- Loeb, N. G. et al. Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. Nat. Geosci. 5, 110-113 (2012).
- Solomon, S. et al. (eds) Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge Univ. Press, Cambridge, 2007).
- Lee, S.-K. et al. Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. *Nat. Geosci.* **8**, 445–449 (2015). Nieves, V., Willis, J. K. & Patzert, W. C. Recent hiatus caused by decadal shift in
- Indo-Pacific heating. *Science* **349**, 532–535 (2015). Chen, X. & Tung, K. K. Varying planetary heat sink led to global-warming slowdown and acceleration. Science 345, 897-903 (2014).
- Thompson, D. W. J. & Solomon, S. Interpretation of recent Southern Hemisphere climate change. Science 296, 895–899 (2002).
- Thornalley, D. J. R. et al. Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years. Nature 556, 227-230 (2018).
- Lozier, M. S. Overturning in the North Atlantic. Annu. Rev. Mar. Sci. 4, 291–315 (2012).
- Robson, J., Ortega, P. & Sutton, R. A reversal of climatic trends in the North Atlantic since 2005. Nat. Geosci. 9, 513-517 (2016).
- Jackson, L. C., Peterson, K. A., Roberts, C. D. & Wood, R. A. Recent slowing of Atlantic overturning circulation as a recovery from earlier strengthening. Nat. Geosci. 9, 518-522 (2016).
- 17. Beckley, B. D. et al. Assessment of the Jason-2 extension to the TOPEX/Pseidon, Jason-1 sea-surface height time series for global mean sea level monitoring. Mar. Geodesy 33, 447-471 (2010).

- 18. Willis, J. K. Can in situ floats and satellite altimeters detect long-term change in Atlantic Ocean overturning? Geophys. Res. Lett. 37, https://doi. org/10.1029/2010GL042372 (2010).
- Yan, X., Zhang, R. & Knutson, T. R. The role of Atlantic overturning circulation in the decent decline of Atlantic major hurricane frequncy. Nat. Commun. 8, (2017)
- 20. Zhang, R. On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. Geophys. Res. Lett. 44, 7865-7875 (2017).
- 21. Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. Nature 556, 191-196 (2018).
- McCarthy, G. D. & Haigh, I. D., Hirschi, J. J.-M., Grist, J. P. & Smeed, D. A. Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. Nature **521**, 508-510 (2015).
- 23. Polyakov, I. V. et al. Arctic Ocean freshwater changes over the past 100 years and their causes. J. Clim. 21, 364-384 (2008).
- 24. Chen, X. & Tung, K. K. Global mean surface temperature variability—space-time perspective from rotated EOFs. Clim. Dyn. https://doi.org/10.1007/s00382-017-3979-0 (2017).
- 25. Wu, Z., Huang, N. E., Wallace, J. M., Smoliak, B. & Chen, X. On the time-varying trend in global-mean surface temperature. Clim. Dyn. 37, 759–773 (2011).
- 26. Zhang, R., Delworth, T. L. & Held, I. M. Can the Atlantic Ocean drive the observed multidecadal variability in Northern Hemisphere mean temperature? Geophys. Res. Lett. 34 1 02709 (2007).
- 27. Tung, K. K. & Zhou, J. Using data to attribute episodes of warming and cooling in instrumental records. Proc. Natl Acad. Sci. USA 110, 2058–2063 (2013)
- 28. Cheng, L. J. & Zhu, J. Artifacts in variations of ocean heat content induced by the observation system changes. Geophys. Res. Lett. 41, 7276-7283 (2014)
- 29. Marshall, J. & Speer, K. Closure of the meridional overturning circulation through Southern Ocean upwelling. Nat. Geosci. 5, 171–180 (2012).
- 30. Yan, X., Zhang, R. & Knutson, T. R. The role of Atlantic Overturning Circulation in the recent decline of Atlantic major hurricane frequency. Nat. Commun. 8, 1695 (2017).

Acknowledgements The research of K.-K.T. is supported by the National Science Foundation, under AGS-1262231 and by the Frederic and Julia Wan Endowed Professorship. X.C. was supported by the National Key Basic Research Program of China under grant 2015CB953900 and by the Natural Science Foundation of China under grants 41330960 and 41776032.

Author contributions K.-K.T. and X.C. undertook the analysis of global ocean temperature and salinity profiles, RAPID observations, and satellite altimetry datasets. K.-K.T. led the draft of this manuscript. X.C. produced all figures. Both authors contributed substantially to the drafting and revision of this manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0320-v

Reprints and permissions information is available at http://www.nature.com/ reprints.

Correspondence and requests for materials should be addressed to K.-K.T. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Updated AMOC indices. We reproduced the unfiltered monthly AMOC indices (Extended Data Fig. 1). Their correlation coefficient with Zhang's unfiltered AMOC fingerprint is listed on the right. All correlations are statistically significant at over 95% confidence level.

AMOC indices in Fig. 3a. Extended Data Fig. 1 shows that all unfiltered AMOC proxies used in Fig. 3a are correlated with Zhang's fingerprint AMOC proxy at over 95% confidence level. Zhang showed²⁰ that in the Geophysical Fluid Dynamics Laboratory model the fingerprint proxy is highly coherent with the model AMOC Index, defined as the zonal integrated maximum Atlantic overturning at 40° N, at decadal and multidecadal scales. This is the reason that the fingerprint is shown smoothed with a 10-year low-pass filter. This fingerprint is calculated using the detrended 400-m subsurface temperature. (It was updated to 2017 by the author with permission to use.)

Our subpolar upper ocean salinity index is defined as the average over 45° – 65° N in the Atlantic basin and integrated over 0–1,500 m. The two undetrended salinity indices shown in Fig. 3 and Extended Data Fig. 1 are from three data sources. The first index is based on ISHII and Scripps. ISHII data have not been updated since 2012 and Scripps data are only available since 2004; they are connected at 2012 when calculating the correlation coefficient with Zhang's fingerprint AMOC proxy. The data source for the second salinity index is from EN4 (version 4.2.1).

The sea-level index was obtained as in ref. 22 by calculating the sea-level difference between the average of a group of linearly detrended, deseasonalized tidegauge measurements south of 35° N and that to the north. It is accumulated in time, shifted to the right by 4.8 years and smoothed with a 7-year lowpass filter.

The subpolar gyre SST index was obtained by 'detrending' the subpolar gyre SST by the subtraction of the global mean SST. It is averaged over the subpolar gyre region, defined by ref. 21 .

Willis' AMOC strength at 41° N was calculated¹⁸ using altimetry SSH measurements and geostrophic approximation for the zonal-mean northward velocity vertically integrated above 1,130 m. It is not detrended or accumulated.

Error bars for data used in Fig. 3. The error bars for the salinity time series used in Fig. 3a are plotted in Extended Data Fig. 2. The uncertainty at each gridpoint is provided by each data source: ISHII, Scripps and EN4. The error bar of the salinity time series at each time is computed as the combination of the gridpoint uncertainty and one standard deviation due to the averaging in space. The uncertainty of the SSH-deduced AMOC strength was given by ref. 18 . The measurement and sampling errors at each time gridpoint were $\pm 12\%$. The uncertainty of tide-gauge data was discussed by ref. 22 , and that of Zhang's fingerprint proxy by ref. 30 . The uncertainty of the global surface temperature data from HadCRUT4.6 was assessed by the data source using 100 ensemble members that span the uncertainty range of the data.

Calculation of warming scenarios. We emphasize that this is not a prediction, but a scenario calculation. In our current climate system, the OHC in the upper 1,500 m of the global oceans increases at the rate of 0.42 W m $^{-2}$, which is approximately the top-of-atmosphere radiative imbalance. Apart from short-term variations of radiative imbalance such as those due to volcanic eruptions, it is reasonable to assume that for the next two decades there will not be an appreciable change in radiative imbalance, barring an unexpected development of carbon sequestration technology.

Scenario I. If the OHC storage below 200 m remains the same (no increases), then the radiative imbalance of the 0.42 W m $^{-2}$ heats only the top 200 m of the global oceans. That is, the increase of OHC in the top 200 m of the oceans is responsible for the increase in the entire 1,500 m of the column. The top 200 m of the global ocean then warms at the rate calculated as: 0.42 W m $^{-2}$ divided by the heat capacity of 200 m of the ocean = 0.23 °C per decade. This is equivalent to that obtained for a 'slab' ocean of 200 m thick.

Scenario 2. As for Scenario 1 except that only the Atlantic and the Southern oceans' heat content below 200 m remain the same for the next two decades. The Pacific and the Indian oceans continue to increase their OHC at the current rate. The warming rate is 70% of that for Scenario 1 because at present the Atlantic and the Southern oceans together are responsible for 70% of the OHC increase in the upper 1,500 m of the oceans. This is probably the more likely scenario because we have argued in the main text that AMOC is likely to remain relatively constant during the next two decades. The subsurface Southern Ocean has been warming since at least 1993, caused by the southward displacement and intensification of the westerly jet, which cannot continue much longer, first because the proposed cause (the ozone hole) has diminished in importance as the ozone hole heals, and second because there is not much more room for the jet's southward displacement. So the increase in warming will probably stop.

Model AMOC and reconciliation with recent observations. Observational results in Fig. 3a show that there was a positive trend from 1993 to 1999, with a small peak in 1996. The rapid rising trend from 1999 to 2005 is statistically significant at the over 95% confidence level. This is seen in all proxies, most clearly in the less

smoothed data (SSH and subpolar salinity). This claim is supported by observation of SSH-deduced AMOC strength, tide-gauges, the subpolar salinity proxy, and also the Zhang fingerprint proxy. (The last proxy, because of 10-year smoothing, does not show the smaller peak in the mid-1990s). A model reanalysis also showed an acceleration prior to 2005 followed by a decline at 26° N, and a peak in the mid-1990s as well as one in 2005 at 45° N¹⁶. AMOC in models is sensitive to resolution and subgrid parameterization 31 , resulting in little consensus among reanalysis (and hindcast) products. With one exception 16 these products do not agree with the RAPID observation at 26° N. The exception is the GloSea5 model, which has a higher, eddy-permitting resolution than previous reanalyses. Supplementary figure 1 of ref. ¹⁶ shows two peaks, one at 1995 and one at 2005. The 1995 peak is slightly higher than the 2005 peak, and is referred to thus in the main text of ref. 16: "The AMOC at 45° N is representative of the changes in the subpolar gyre, with the AMOC decreasing from a maximum in the mid-1990s, followed by a slight increase (Fig. 1d)". The peak in 2005 was not mentioned. However, the result on the 1995 peak should be treated with care, as the authors themselves stated in the supplementary information of ref. ¹⁶: "It is likely that there will be a period of spinup, where the deep ocean (where there are few observational constraints) adjusts, which may explain the divergence in trend. Hence we disregard the first few years of each experiment. There is also a shock in 1992 when the altimeter data is introduced, which may contribute to the increase in AMOC strength between 1989 and 1995. Hence we choose the period to analyse starting from January 1995, and join the two analyses in January 2002." The relative magnitude of the 1995 peak and the 2005 peak may be unreliable as it was obtained by joining two reanalyses, one starting from 1989 and one from 1995 with "divergence in trend" 16.

The observed SSH data since 1992 can be used to deduce AMOC strength using geostrophic approximation, bypassing the problems of shock and subsequent adjustment when the same SSH data were introduced in model assimilation.

SST changes during different phases of AMOC. The upper branch of the climatological AMOC brings warm and saline surface water from the subtropical North Atlantic to its subpolar latitudes. When the overturning is stronger, more of this warm water is found in the subpolar northern latitudes. In the Southern Hemisphere, more of the cold water from the region of the Antarctic Circumpolar Current is brought northward into the Southern subtropics. Consequently a characteristic signature in the Atlantic SST is an opposite-signed multidecadal anomaly, with warming to the north and smaller cooling to the south when the overturning is stronger (AMOC+), and the reverse pattern when it is weaker (AMOC-) (Extended Data Fig. 5a, b). This ocean-induced SST variability is centered in the subpolar North Atlantic²⁰. The observed tendency during the last two subphases of the AMOC is as expected (Extended Data Fig. 5c, d): As AMOC slows after 2005, the SST tends towards a cooler North Atlantic and warmer subtropics. Accompanying the strong cooling in the subpolar gyre is an interesting intense warming after 2005 in the northwest Atlantic, centered in the Gulf of Main, which was recently simulated in a high-resolution climate model³² as due to the northward displacement of the Gulf Stream when AMOC slows. The inverse relationship between Gulf Stream's northward displacement and AMOC strength was found⁶ to be caused by the Labrador Current retreat and the bottom vortex stretching³³. AMO. In long coupled atmosphere-ocean model runs under preindustrial conditions (without increasing greenhouse gases) the AMO is the SST manifestation of AMOC variations, and the two time series are approximately in phase¹⁹. The definition of AMO in ref. 19 is the mean of Atlantic SST north of 45° N, which may lead the subtropical SST anomaly by two years. A more traditional definition of AMO is the mean Atlantic SST north of the Equator³⁴, with an approximately one-year phase difference. It has been shown²⁴, using the space-time perspective of rotated empirical orthogonal function analysis, that the AMO is mainly responsible for the observed global mean surface temperature variation on multidecadal timescales. The two are in phase during the industrial era. Since the AMOC and the global mean surface temperature variation are not in phase (as shown in Fig. 3), it follows that during the industrial era, AMOC and AMO are off phase, possibly by a quarter cycle, although AMOC's time series is too short for an accurate determination of the phase information.

During the positive phase of AMO, SST is warm in the North Atlantic and surrounding continents. Therefore, Northern Hemisphere mean surface temperature is warm during the positive phase and cool during the negative phase of the AMO. Using multiproxy data in the Northern Hemisphere the AMO time series can be extended back several hundred years³⁵. The longest instrumental temperature record exists in central England, and it was used²⁷ to reconstruct the AMO timeseries back to the Little Ice Ages. An even longer record of ice cores in Greenland, in the northern Atlantic, exists, and a statistically significant at the over 95% confidence level AMO signal can be found³⁶ extending back to 800 AD that is coherent with the instrumental record of central England²⁷ during their overlapping period. It appears that AMO is a recurrent phenomenon of period around 65–70 years and that it is robust in the preindustrial era, with the Atlantic and the surrounding areas warm during the positive phase and cold during the negative phase. From climate

model preindustrial control runs, it seems that AMO is a surface manifestation of AMOC variation. Furthermore, based on palaeoclimate evidence of cold events when AMOC slows down abruptly, a common perception is that a slowdown in AMOC would lead to a cold Northern Hemisphere. The mechanism relies on the dominant role of AMOC (and its Gulf Stream) in horizontally transporting surface heat from the tropics to the mid- and high-latitude Atlantic, where it releases some heat to the cold atmosphere before sinking in the subpolar Atlantic. The heat released to the atmosphere makes Europe warmer (when wind blows in that direction) than it should be for its latitude.

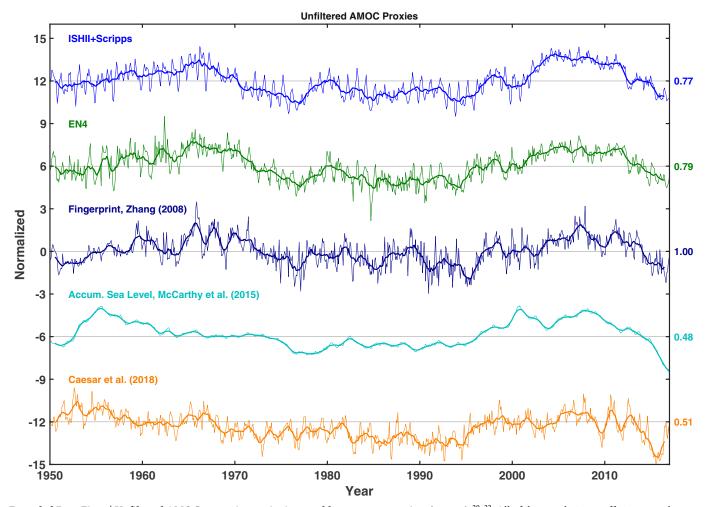
Calculating SSH* from altimetry data. SSH* is SSH with its global mean subtracted. SSH contains both the thermosteric part (due to thermal expansion of the entire water column) and the ocean water mass addition that is due to melting land ice. It is known that the ocean will adjust to any change in ocean mass rapidly through the propagation of gravity waves, and will reach a new equilibrium globally within a couple of months³⁷. Therefore, the subtraction of the global mean largely removes the mass contribution from SSH.

Data availability. The datasets used in this study are all publicly available. They are: (1) ISHII data version 6.13, the objectively analysed subsurface temperature and salinity at 24 levels in the upper 1,500 m during 1945–2012 (http://rda.ucar. edu/datasets/ds285.3/); (2) Scripps gridded Argo data, objectively analysed subsurface temperature and salinity at 58 levels in the upper 1,950 m since 2004 (http://www.argo.ucsd.edu/Gridded_fields.html), which is based on Argo data collected and made freely available by the international Argo project and the national programmes that contribute to it; Argo float data and metadata are available from the Argo Global Data Assembly Centre (https://doi.org/10.17882/42182); (3) EN4 data version 4.2.1, objectively analysed subsurface temperature and salinity at 42 levels in the upper 5,350 m since 1900 (https://www.metoffice.gov.uk/hadobs/

en4/download-en4-2-1.html); (4) Sea surface height based on satellite altimetry from the Archiving, Validation, and Interpretation of Satellite Oceanographic Data (AVISO) (https://www.aviso.altimetry.fr/en/data.html); (5) Tide gauge records from the Permanent Service of Mean Sea Level (PSMSL) (http://www.psmsl.org); (6) Extended Reconstructed Sea Surface Temperature (ERSST, version 3b) (http://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v3b/netcdf); (7) RAPID AMOC at 26.5° N (http://www.rapid.ac.uk/rapidmoc/rapid_data/); (8) Ref. ¹⁸, updated by the author (ftp://oceans-ftp.jpl.nasa.gov/pub/jwillis/AMOC/2016/).

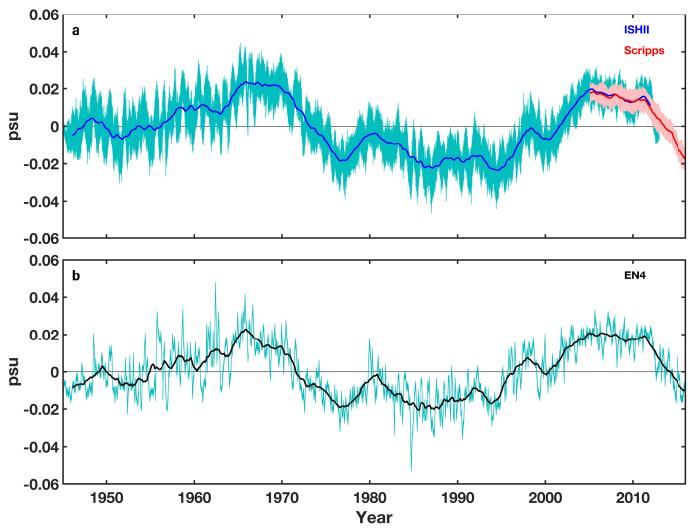
Code availability. Scripts for analysing the data are available from the corresponding author upon reasonable request.

- 31. Danabasoglu, G. et al. North Atlatnic simulations in cordinated ocean-ice reference experiments phase II (CORE-II). Part II: interannual to decadal variability. *Ocean Model.* **97**, 65–90 (2016).
- Saba, V. S. et al. Enhanced warming of the Northwest Atlantic Ocean under climate change. J. Geophys. Res. Oceans 121, 118–132 (2016).
- Zhang, R. & Vallis, G. K. The role of bottom vortex stretching on the path of the North Atlantic Western Boundary Current and on the North Recirculation Gyre. J. Phys. Oceanogr. 37, 2053–2080 (2007).
- Enfield, D. B., Mestas-Nunez, A. M. & Trimble, P. J. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U. S. Geophys. Res. Lett. 28, 2077–2080 (2001).
- 35. Delworth, T. L. & Mann, M. E. Observed and simulated multidecadal variability in the Northern Hemisphere. *Clim. Dyn.* **16**, 661–676 (2000).
- Zhou, J., Tung, K. K. & Li, K.-F. Multidecadal variability in the Greenland ice-core records obtained using Intrinsic Time-scale Decomposition. *Clim. Dyn.* 47, 739–752 (2016).
- Lorbacher, K., Marsland, S. J., Church, J. A., Griffies, S. M. & Stammer, D. Rapid barotropic sea level rise from ice sheet melting. *J. Geophys. Res.* 117, (2012).

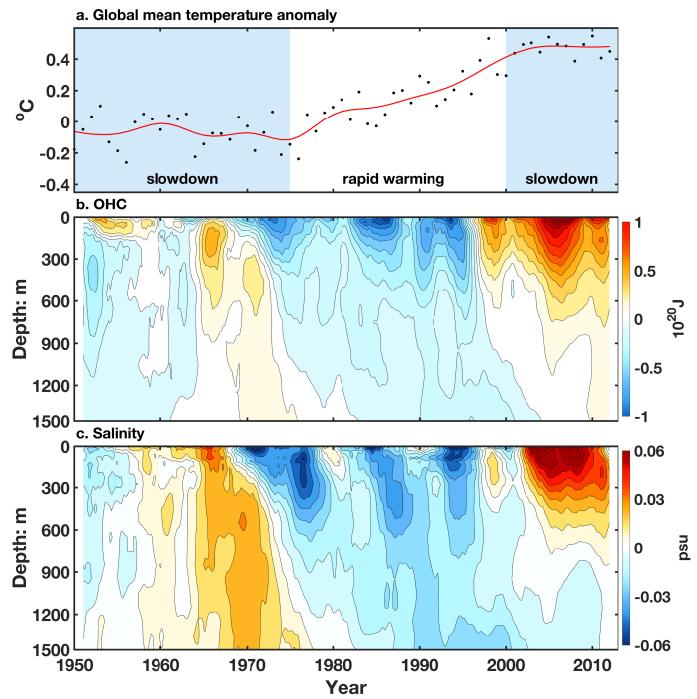


Extended Data Fig. 1 | **Unfiltered AMOC proxy time series in monthly resolution.** The thick solid lines are 13-month running means. The numbers to the right of each time series show the correlation coefficient with the unfiltered AMOC subsurface temperature fingerprint of Zhang.

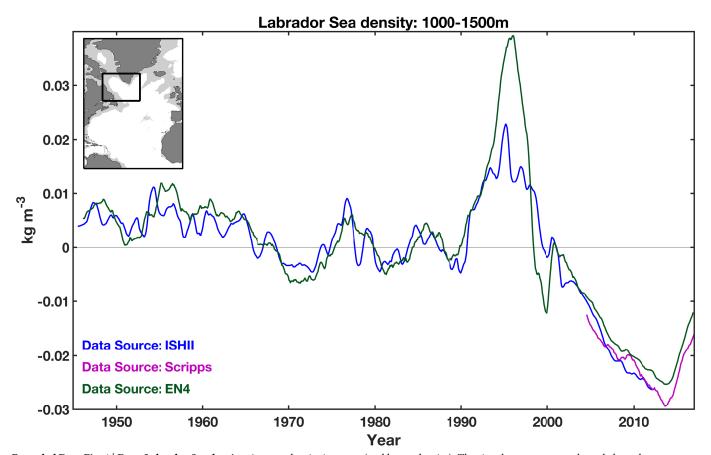
Data are taken from refs $^{20-22}$. All of the correlation coefficients are above 95% confidence level. The accumulated sea-level index is shifted to the right by 4.8 years in this figure. Without the time shift, its correlation with the AMOC proxy is practically zero (r= 0.06).



Extended Data Fig. 2 | **Error bars for the three salinity time series shown in Fig. 1.** The colour lines are monthly values of uncertainty, superimposed on the 13-month means of the time series. psu, practical salinity units.

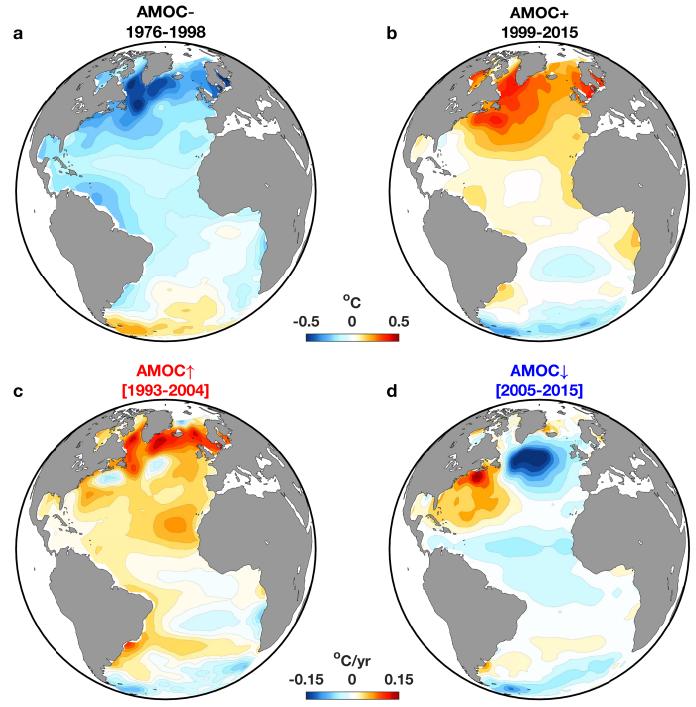


Extended Data Fig. 3 | **Coincidence of the three AMOC phases with global warming slowdown and acceleration. a**, Global mean surface temperature. **b**, OHC north of 45° N in the Atlantic. **c**, Salinity north of 45° N in the Atlantic.

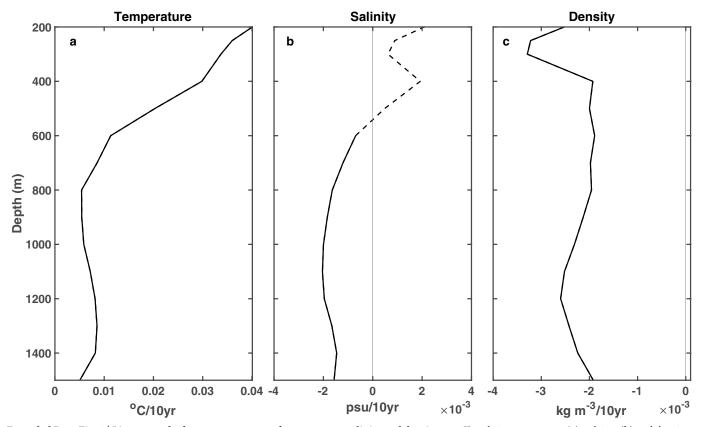


Extended Data Fig. 4 | Deep Labrador Sea density: Average density in the 1,000-1,500 m layer of the Labrador Sea, regionally averaged over the ocean area shown in the inset, from the three data sources given. A leading signal for stronger AMOC is the increased deep Labrador Sea salinity

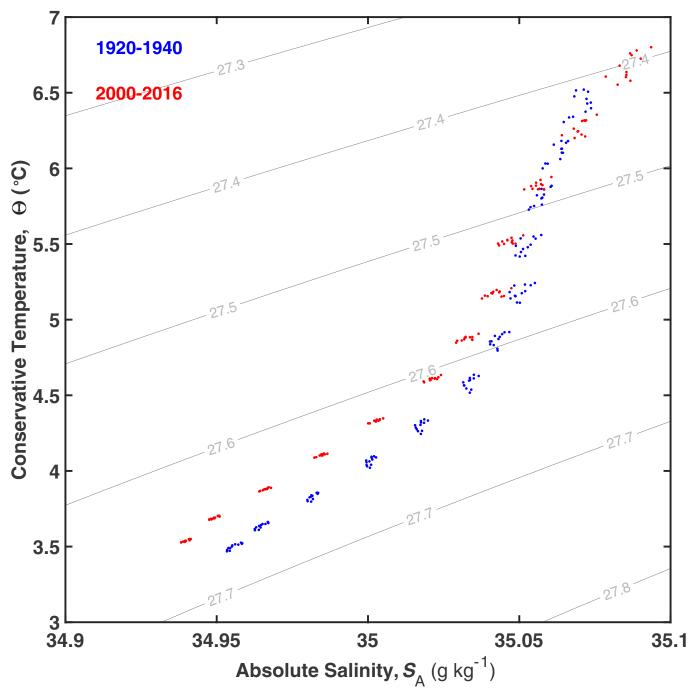
(and hence density). The signal propagates southward along the western boundary at depth, changing the cross-basin zonal gradient, and hence the geostrophic southward velocity 13 . The return flow then strengthens the upper branch of AMOC with a lag of 7–10 years 15,16 .



Extended Data Fig. 5 | **SST patterns during different AMOC phases. a**, When AMOC is below climatology. **b**, When AMOC is above climatology, SST detrended. **c**, SST linear trend when AMOC is increasing. **d**, When AMOC is decreasing.



Extended Data Fig. 6 | Linear trends, from 1950 to 2017, of temperature, salinity and density. a-c, Trends in temperature (a), salinity (b) and density (c) as a function of depth. Solid curves indicate where the trend is statistically significant at 95% confidence level.



Extended Data Fig. 7 | **Temperature–salinity diagram.** The subpolar Atlantic Ocean (45° – 65° N) for each depth between 300 m and 1,500 m for the two periods, with the mean of 2000–2016 in red and the mean of 1920–

1940 in blue. The dots shown are the five winter month values (NDJFM). At these depths the seasonal cycle is very small 38 .



An inverse latitudinal gradient in speciation rate for marine fishes

Daniel L. Rabosky^{1,10}*, Jonathan Chang^{2,10}, Pascal O. Title^{1,10}, Peter F. Cowman^{3,4}, Lauren Sallan⁵, Matt Friedman⁶, Kristin Kaschner⁷, Cristina Garilao⁸, Thomas J. Near³, Marta Coll⁹ & Michael E. Alfaro^{2,10}

Far more species of organisms are found in the tropics than in temperate and polar regions, but the evolutionary and ecological causes of this pattern remain controversial^{1,2}. Tropical marine fish communities are much more diverse than cold-water fish communities found at higher latitudes^{3,4}, and several explanations for this latitudinal diversity gradient propose that warm reef environments serve as evolutionary 'hotspots' for species formation⁵⁻⁸. Here we test the relationship between latitude, species richness and speciation rate across marine fishes. We assembled a time-calibrated phylogeny of all ray-finned fishes (31,526 tips, of which 11,638 had genetic data) and used this framework to describe the spatial dynamics of speciation in the marine realm. We show that the fastest rates of speciation occur in species-poor regions outside the tropics, and that high-latitude fish lineages form new species at much faster rates than their tropical counterparts. High rates of speciation occur in geographical regions that are characterized by low surface temperatures and high endemism. Our results reject a broad class of mechanisms under which the tropics serve as an evolutionary cradle for marine fish diversity and raise new questions about why the coldest oceans on Earth are present-day hotspots of species formation.

The steep decline in species richness from the equator to the poles is one of the most general large-scale patterns in biology ^{9,10} and has existed in its general form for more than 30 million years ¹¹. Many proposed mechanisms for this latitudinal diversity gradient (LDG) explain high tropical diversity as the outcome of faster rates of species origination: the tropics are an evolutionary cradle for new species, and the gradient reflects—at least in part—lower rates of species formation in regions outside the tropics ^{1,12}. Studies on fossil mollusks ¹², plankton ¹³ and corals ⁵ support the hypothesis that rates of marine species formation are faster in the tropics than at higher latitudes.

We tested whether latitudinal variation in the rate of speciation can explain the LDG in marine fish diversity by reconstructing speciation rates across fishes and analysing them in a geographical context. We focused explicitly on recent speciation rates^{2,14,15}, because extinction reduces our ability to infer rates deep in the past¹⁶. We also ignored phylogenetic estimates of extinction rates, given the unreliable nature of these parameters in phylogenetic diversification models¹⁷. If speciation rates are controlled by energy—perhaps owing to accelerated chemical reactions, life histories or mutation rates^{18,19}—then we should observe a footprint of rapid speciation in the distribution of recent speciation times for tropical taxa.

We assembled a distribution of all-taxon assembled (ATA) time-calibrated phylogenetic trees of ray-finned fishes (31,526 species). The ATA phylogenies include 11,638 species with genetic data (5,231 marine species); the remaining 19,888 species that did not have genetic data were placed using stochastic polytomy resolution (Methods) to

generate taxonomically consistent resolutions of all taxa without genetic data under a conservative constant-rate birth-death process. The ATA trees were time-calibrated using a database of 139 fossil taxa (Extended Data Fig. 1 and Supplementary Information). We estimated or compiled geographic ranges for the majority of known marine species, including all species with genetic data. We estimated speciation rates across the phylogenies using BAMM²⁰, a Bayesian framework for reconstructing complex evolutionary dynamics from phylogenetic trees, and DR, a summary statistic that infers recent speciation rates for all tips in the phylogeny without requiring a formal parametric inference model²¹. We denote these two analyses of speciation rates as $\lambda_{\rm BAMM}$ and $\lambda_{\rm DR}$, respectively. The $\lambda_{\rm BAMM}$ and $\lambda_{\rm DR}$ rates include substantial historical information and are best interpreted as the rate of lineage splitting averaged across the past 10–20 million years (Myr)²; units for speciation presented here are per-lineage rates in units of lineages per Myr. We also computed a simple interval-based measure of speciation rate for a series of path intervals from 0.25 Myr to 50 Myr before present²², providing a window of reliability for λ_{BAMM} and λ_{DR} . Estimates of λ_{DR} were computed across the distribution of ATA phylogenies, thus generating rate estimates conditional on the uncertainty in placements of taxa without genetic data. $\lambda_{\rm BAMM}$ was estimated from the primary dated tree including all taxa with genetic data (n = 11,638), and incomplete sampling was incorporated by using family-specific sampling fractions.

Consistent with previous studies^{3,4}, we find a strong LDG in marine fish diversity, with an extreme richness peak in the Coral Triangle of the tropical Indo-Pacific Ocean (Fig. 1a). However, analysis of per-cell mean speciation rates reveals a notable inverse relationship between the rate of species formation and latitude (Fig. 1b-e). Mean speciation rate for cell assemblages from tropical regions ($<23.5^{\circ}$; n = 6,698cells) was $\lambda_{\rm BAMM}$ = 0.08 ($\lambda_{\rm DR}$ = 0.11) and the corresponding rate for high-latitude (>45°; n = 4,347 cells) assemblages was $\lambda_{\rm BAMM}$ = 0.14 ($\lambda_{DR} = 0.16$). These rate differences are substantially greater when comparing more species-rich assemblages from continental shelf and slope regions: shallow (mean depth <2,000 m) tropical cells have $\lambda_{\rm BAMM} = 0.08$ ($\lambda_{\rm DR} = 0.11$; n = 833), whereas corresponding highlatitude cells have $\lambda_{\text{BAMM}} = 0.18$ ($\lambda_{\text{DR}} = 0.22$; n = 1,182). We computed means for 232 marine biogeographic ecoregions—encompassing the Earth's shallow and coastal regions—and used spatial simultaneous autoregressive (SAR) models with breakpoints to assess the relationship between latitudinal position and speciation rate. Regardless of how regional mean rates are computed, all SAR models have highly significant effects of latitude on speciation rate (P < 0.001; n = 232regions; Extended Data Fig. 2). In general, for latitudes greater than 25° north or south, each ten-degree increase in latitude increases the assemblage-wide speciation rate by approximately 0.025 lineages per Myr. However, speciation rate is effectively decoupled from latitude

¹Museum of Zoology, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. ²Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. ³Peabody Museum of Natural History and Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. ⁴ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia. ⁵Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, PA, USA. ⁶Museum of Paleontology and Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA. ⁷Department of Biometry and Environmental System Analysis, Albert-Ludwigs-University of Freiburg, Freiburg, Germany. ⁸GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany. ⁹Instituto de Ciencias de Mar, Spanish National Research Council (ICM-CSIC), Barcelona, Spain. ¹⁰These authors contributed equally: Daniel L. Rabosky, Jonathan Chang, Pascal O. Title, Michael E. Alfaro. *e-mail: drabosky@umich.edu

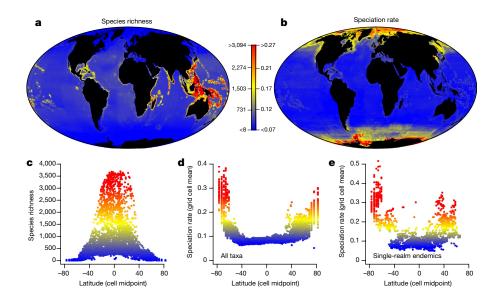


Fig. 1 | Latitudinal gradient in species diversity and speciation rate in marine fishes. a, b, Mean species richness (a) and speciation rate (b) for marine fish assemblages at the global scale. c, d, Marginal distributions of richness (c) and speciation rate (d) with respect to latitude (n=16,150), with cell colours corresponding to scale bars in a, b. e, Mean speciation rates for endemic taxa only (n=2,698). Results shown for $\lambda_{\rm BAMM}$ but similar results are obtained for $\lambda_{\rm DR}$ (Extended Data Figs. 2, 3). Grid cell size is 150×150 km for all panels.

for tropical and subtropical regions (Extended Data Fig. 2h). General results reported here are robust across all of the measures of speciation rate and associated weighting schemes that we considered (Extended Data Figs. 2, 3).

Speciation rate is strongly and negatively associated with both species richness (Fig. 2a) and annual sea surface temperature (Fig. 2b), although sea surface temperature is highly correlated with latitude (r = -0.95across 16,150 grid cells). Regional assemblages of fishes with the fastest rates of speciation occur at the highest latitudes and are characterized by cold surface temperatures (Fig. 2 and Extended Data Fig. 4). The south polar seas, dominated by the in situ radiation of highly specialized and geographically restricted icefishes and their relatives²³, are characterized by the fastest overall rates of species formation of any marine region on Earth. Continental shelf and slope assemblages from the Southern Ocean surrounding Antarctica have mean speciation rates of $\lambda_{\rm BAMM} = 0.27$ and $\lambda_{\rm DR} = 0.26$ (n = 179 cells); these rates substantially exceed those observed for the Coral Triangle ($\lambda_{\text{BAMM}} = 0.08$ and $\lambda_{\rm DR} = 0.11$; n = 220 cells), despite a mean 62-fold difference in per-cell species richness for these regions. Assemblages from the Arctic also have high speciation rates ($\lambda_{\text{BAMM}} = 0.17$ and $\lambda_{\text{DR}} = 0.24$; n = 511 cells), despite little overlap between the clades that comprise the northern and southern polar faunas²⁴. There is a strong positive relationship between several analyses of regional endemism and assemblage-wide speciation rate (Fig. 2c and Extended Data Fig. 4e; n = 60 regions). The correlation between λ and endemism is high overall (λ_{BAMM} , r = 0.81; λ_{DR} , r = 0.79). The Mediterranean Sea is a clear outlier with respect to this overall pattern, combining high endemism with relatively low speciation (Fig. 2c). This suggests that the factors contributing to endemism per se are not necessarily those that promote fast speciation.

As an alternative to the analysis of mean speciation rates by grid cell and biogeographical region (Figs. 1, 2), we analysed λ_{BAMM} and λ_{DR} for individual fish species with respect to their latitudinal midpoint. High-latitude fish clades are characterized by rapid speciation relative to low-latitude and reef-associated clades, and there is a strong relationship between the centroid midpoint of the geographic range for each species and its estimated rate of species formation (Fig. 3 (inset) and Extended Data Fig. 5). We formally tested the relationship between latitudinal midpoint and speciation rate using several methods that are robust to model misspecification and phylogenetic pseudoreplication^{25,26}. The correlation between absolute latitudinal midpoint and $\lambda_{\rm DR}$ is 0.27 (P < 0.001); similar results are obtained for $\lambda_{\rm BAMM}$ and latitude (r = 0.3; P = 0.006). Across a range of latitudinal thresholds, we find a highly significant difference in speciation rate for high- and low-latitude fishes (P < 0.001 across all thresholds), and coldtemperate and polar lineages speciating approximately twice as fast as the average low-latitude lineage (Extended Data Table 1).

Species with latitudinal midpoints in the tropics (23.5° S to 23.5° N; n=3,461) have mean speciation rates of $\lambda_{\rm BAMM}=0.09$ and $\lambda_{\rm DR}=0.12$. By contrast, species with latitudinal midpoints greater than 45° N or 45° S (n=574) have $\lambda_{\rm BAMM}=0.20$ and $\lambda_{\rm DR}=0.25$. These rates are even more extreme for subpolar and polar taxa: across fishes in our dataset with latitudinal midpoints greater than 60° (n=122), mean speciation rates were $\lambda_{\rm BAMM}=0.29$ and $\lambda_{\rm DR}=0.35$. Interval-based estimates of speciation rate²² indicate that the overall tropical–temperate–polar gradient that we report here has been present for millions of years, extending back in time at least the Miocene/Pliocene boundary (Extended Data Fig. 6).

Reef-associated clades, which comprise a substantial fraction of the tropical diversity peak, are not characterized by exceptional rates of species formation. Three of the largest such clades—the wrasses, damselfishes and gobies—collectively account for approximately 3,000 species, yet have low to moderate rates of speciation estimated using BAMM (wrasses: $\lambda_{\text{BAMM}} = 0.10$; gobies: $\lambda_{\text{BAMM}} = 0.07$; damselfishes: $\lambda_{\text{BAMM}} = 0.12$) and DR (wrasses: $\lambda_{\text{DR}} = 0.12$; gobies: $\lambda_{\text{DR}} = 0.10$; damselfishes: $\lambda_{DR} = 0.14$). By contrast, temperate and polar fish faunas are dominated by members of multiple clades that have exceptionally high rates of species formation (Fig. 3), including snailfishes, eelpouts, Sebastes rockfishes and Antarctic notothens (icefishes and allied species). These coldwater taxa are characterized by speciation rates that exceed 0.26 (λ_{BAMM}) and 0.34 (λ_{DR}). With the possible exception of gobies, we find little evidence for early bursts of speciation during the radiations of major tropical and reef-associated clades across the past 20-60 Myr (Extended Data Fig. 6). We note that 79.7% of marine speciation events in our ATA phylogenies are inferred to have occurred after the Oligocene/Miocene boundary, suggesting that the timescales over which we have estimated speciation rates are relevant to the origin and maintenance of modern LDG in marine fishes.

An alternative explanation for the global gradient in speciation rates that we report involves environmental or biogeographical filtering on traits that are also associated with rapid speciation. For example, perhaps speciation rates are most rapid for fishes that inhabit cold and dark bathyal or abyssal regions; physiological adaptations for life in those environments might predispose these lineages towards disproportionate representation in high-latitude communities. This hypothesis predicts that deep-sea lineages should speciate more rapidly than shallow lineages, regardless of latitude. However, mean rates for high-latitude (>45°) deep-sea fishes are much faster than for low-latitude (<45°) deep-sea species (high latitude: $\lambda_{\rm BAMM} = 0.29$, $\lambda_{\rm DR} = 0.37$, n = 75; low latitude: $\lambda_{\rm BAMM} = 0.15$, $\lambda_{\rm DR} = 0.15$, n = 218). Across all deep-sea fishes represented in our dataset (n = 293), there is a strong positive correlation between absolute latitudinal midpoint and speciation rate (r = 0.50; P < 0.001). There is no effect of depth classification on speciation rate

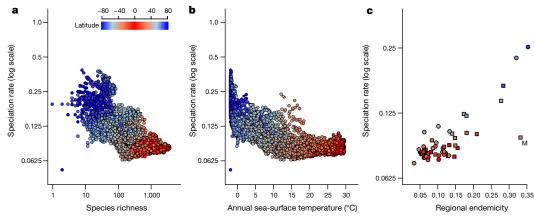


Fig. 2 | Species richness, temperature and speciation rate in marine fishes for individual grid cells. a, Negative relationship between species richness and mean speciation rate ($\lambda_{\rm BAMM}$) for individual grid cells (n=16,150). b, Negative relationship between mean annual sea-surface temperature and mean speciation rate for cells. c, Positive relationship between regional endemism and mean speciation rate for all species occurring in a particular biogeographical province (n=60 biogeographical

provinces). Squares and circles denote provinces with latitudinal midpoints north and south of the equator, respectively; cell colours denote latitude. Point labelled 'M' in the lower right of ${\bf c}$ is the Mediterranean Sea, which is characterized by high endemism and low speciation rate. Nearly identical results are obtained for $\lambda_{\rm DR}$ and for BAMM analyses that assume time constancy within rate regimes (Extended Data Fig. 4).

for tropical fishes (P > 0.25 across all classification schemes; Extended Data Fig. 7a). A secondary prediction of the filtering hypothesis is that high-rate, high-latitude clades should be nested within high-rate

tropical or deepwater clades. We tested this hypothesis for perciform fishes, which account for 66.3% of high-latitude fishes (Extended Data Fig. 7b). Perciformes include four of the most-rapidly speciating

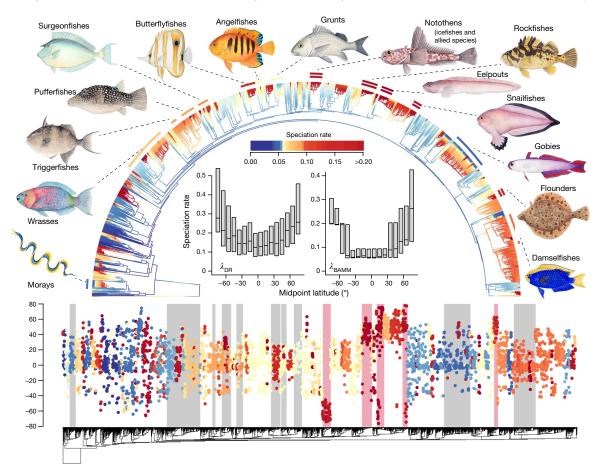


Fig. 3 | Latitudinal gradient in per-taxon speciation rate for marine fishes. Top, BAMM-estimated speciation rates across phylogenetic tree of 5,223 marine fishes for which genetic and geographic range data were available. Iconic coral reef clades are indicated with single arc segments; double segments denote high-latitude lineages that drive the overall fast speciation rate for temperate and polar fishes. Inset box plots show the median and interquartile range in distribution of rates ($\lambda_{\rm DR}$ and $\lambda_{\rm BAMM}$) for individual taxa with respect to the centroid midpoint of their

latitudinal distribution, with species values binned in 10° increments. Bottom, phylogenetic niche conservatism in marine fish lineages as reflected by the geographical distribution of latitudinal midpoints; each point is the centroid midpoint of an individual species, and colours reflect corresponding $\lambda_{\rm BAMM}$ estimates. Clades denoted with pink polygons are dominant high-latitude fish clades; grey polygons are predominantly reef-associated clades. The fish images were created by J. Johnson.

major clades of marine fishes (Notothenioids, Sebastidae, Zoarcidae and Liparidae), but these high-latitude clades are either nested within other high-latitude clades or within largely tropical clades that have low speciation rates (Extended Data Fig. 7c). The overall latitudinal gradient in speciation rate is thus unlikely to be explained by filtering on deepwater clades with rapid speciation rates into high-latitude biogeographical provinces.

We performed a complementary set of analyses based only on primary occurrence records from museum databases and other sources (see Methods). These estimates of species ranges yield highly congruent results (Extended Data Fig. 8). Our results are not conditional on a specific parametric model for inference; the terminal branch lengths themselves are strongly associated with latitude (Extended Data Fig. 9), indicating that few assumptions are required to obtain the results presented here. Furthermore, these results cannot be explained by variation in the completeness of taxonomic sampling with respect to latitude or by alternative reconstructions of geographic range (Extended Data Figs. 8, 9).

Our results reject the hypothesis that rapid speciation explains the spectacular diversity of tropical marine shallow-water fishes and reveal that, paradoxically, speciation rates are fastest in the geographical regions with the lowest species richness. Several evolutionary explanations for the LDG propose that fundamental relationships between energy and speciation rate control the accumulation of biodiversity over time^{18,19}, and it has been said that the tropics are more diverse because 'the Red Queen runs faster when she is hot'²⁷. For the marine fish species that were studied here—and for many terrestrial vertebrates^{2,21}—there is no evidence to support these biophysical linkages between energy, metabolism and speciation. Faster speciation contributes to total species richness in some island and freshwater lacustrine systems^{2,28}, but for larger biogeographical provinces including the marine realms considered in the present study—it is increasingly unlikely that speciation rate variation is the primary cause of diversity gradients^{2,29}. Whether the rapid speciation that we have documented in Earth's cold oceanic provinces reflects a recent and ongoing expansion of marine diversity is a key frontier for future research on the LDG in marine organisms.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0273-1

Received: 2 July 2017; Accepted: 29 May 2018; Published online 4 July 2018.

- Mittelbach, G. G. et al. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol. Lett. 10, 315–331 (2007).
- Schluter, D. & Pennell, M. W. Speciation gradients and the distribution of biodiversity. Nature 546, 48–55 (2017).
- Tittensor, D. P. et al. Global patterns and predictors of marine biodiversity across taxa. Nature 466, 1098–1101 (2010).
- Stuart-Smith, R. D. et al. Integrating abundance and functional traits reveals new global hotspots of fish diversity. Nature 501, 539–542 (2013).
- Kiessling, W., Simpson, C. & Foote, M. Reefs as cradles of evolution and sources of biodiversity in the Phanerozoic. Science 327, 196–198 (2010).
- Alfaro, M. E., Santini, F. & Brock, C. D. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). Evolution 61, 2104–2126 (2007).
- Cowman, P. F. & Bellwood, D. R. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. J. Evol. Biol. 24, 2543–2562 (2011).
- Siqueira, A. C., Oliveira-Santos, L. G. R., Cowman, P. F. & Floeter, S. R. Evolutionary processes underlying latitudinal differences in reef fish biodiversity. Glob. Ecol. Biogeogr. 25, 1466–1476 (2016).
- Hillebrand, H. On the generality of the latitudinal diversity gradient. Am. Nat. 163, 192–211 (2004).
- 10. MacArthur, R. H. Geographical Ecology (Princeton Univ. Press, Princeton, 1972).

- 11. Mannion, P. D., Upchurch, P., Benson, R. B. J. & Goswami, A. The latitudinal biodiversity gradient through deep time. *Trends Ecol. Evol.* **29**, 42–50 (2014).
- Jablonski, D., Roy, K. & Valentine, J. W. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. Science 314, 102–106 (2006).
- Allen, A. P. & Gillooly, J. F. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. Ecol. Lett. 9, 947–954 (2006).
- Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315, 1574–1576 (2007).
- Rabosky, D. L., Title, P. O. & Huang, H. Minimal effects of latitude on present-day speciation rates in New World birds. Proc. R. Soc. Lond. B 282, 20142889 (2015).
- Liow, L. H., Quental, T. B. & Marshall, C. R. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Syst. Biol. 59, 646–659 (2010).
- Davis, M. P., Midford, P. E. & Maddison, W. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* 13, 38 (2013).
- Rohde, K. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65, 514–527 (1992).
- Allen, A. P., Brown, J. H. & Gillooly, J. F. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. Science 297, 1545–1548 (2002).
 Rabosky, D. L., Mitchell, J. S. & Chang, J. Is BAMM flawed? Theoretical and
- Rabosky, D. L., Mitchell, J. S. & Chang, J. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. Syst. Biol. 66, 477–498 (2017).
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* 491, 444–448 (2012).
- Freckleton, R. P., Phillimore, A. B. & Pagel, M. Relating traits to diversification: a simple test. Am. Nat. 172, 102–115 (2008).
- Near, T. J. et al. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Natl Acad. Sci. USA 109, 3434–3439 (2012).
- Eastman, J. T. Comparison of the Antarctic and Arctic fish faunas. Cybium 21, 335–352 (1997).
- Harvey, M. G. & Rabosky, D. L. Continuous traits and speciation rates: alternatives to state-dependent diversification models. *Methods Ecol. Evol.* 9, 984–993 (2018).
- Rabosky, D. L. & Huang, H. A robust semi-parametric test for detecting trait-dependent diversification. Syst. Biol. 65, 181–193 (2016).
- 27. Brown, J. H. Why are there so many species in the tropics? *J. Biogeogr.* **41**, 8–22 (2014).
- Wagner, C. E., Harmon, L. J. & Seehausen, O. Cichlid species—area relationships are shaped by adaptive radiations that scale with area. *Ecol. Lett.* 17, 583–592 (2014).
- Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. Nature 555, 246–250 (2018).

Acknowledgements We thank M. Grundler for statistical and coding advice, and M. Venzon and A. Noonan for assistance with dataset assembly. We are grateful to the many institutions that curate the primary biodiversity data that underlie several of our analyses (see Supplementary Table 6). This research was carried out using computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. This work was supported in part by NSF grant DEB-1256330 (D.L.R.), an NSF DDIG grant to J.C. (DEB-1601830), an Encyclopedia of Life Rubenstein Fellowship to J.C. (EOL-33066-13) and by a Fellowship from the David and Lucile Packard Foundation (D.L.R.). P.F.C. was funded by a Gaylord Donnelley Postdoctoral Environment Fellowship (Yale) and through the ARC Centre of Excellence for Coral Reef Studies. We thank J. Johnson for creating the fish images in Fig. 3 and Extended Data Fig. 7.

Reviewer information *Nature* thanks O. Bininda-Emonds, O. Seehausen and the other anonymous reviewer(s) for their contribution to the peer review of this work

Author contributions D.L.R. and M.E.A. designed the study. D.L.R. drafted the paper with substantial input from P.O.T., M.E.A. and J.C. J.C., P.O.T., M.E.A., P.F.C., L.S., M.F., K.K., C.G., T.J.N., M.C. and D.L.R. contributed data. J.C., P.O.T. and D.L.R. developed methods, and P.O.T. and J.C. developed pipelines for data processing and analysis. D.L.R., P.O.T., J.C. and M.E.A. analysed data. All authors contributed to interpretation and discussion of results. Authorship order for P.O.T. and J.C. was determined by coin toss.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0273-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0273-1.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.L.R. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Matrix assembly. We used the PHLAWD pipeline³⁰ to generate a 27-gene multilocus alignment for ray-finned fishes (Supplementary Information). Guide alignments were constructed using data from recently published studies of higher-level actinopterygian relationships^{31,32}. Guide alignments also included new sequences for 442 species of actinopterygians (Supplementary Table 2; see 'Data availability') generated using a standardized phylogenetic workflow for fishes³². PHLAWD produced a preliminary alignment of 15,606 species. We performed a series of curation steps including BLAST searches of each sequence back to GenBank to identify taxonomically misassigned species, taxonomic name reconciliation against the California Academy of Sciences taxonomy, duplicate species detection and visual identification of poorly aligned sequences (Supplementary Information). We removed rogue sequences using the RogueNaRok searches³³ and performed preliminary tree searches in RAxML to identify and remove sequences with pathologically long branches due to misalignment. After curation of the raw alignment, our final alignment contained 11,638 species. We used PartitionFinder³⁴ to identify a model of sequence evolution for multigene alignment and RAxML to find the maximum likelihood topology and calculate Shimodaira-Hasegawa-like support values³⁵ (Supplementary Information).

Divergence time analysis. We surveyed the palaeontological literature and museum catalogues to assemble our actinopterygian fossil calibration set (139 early occurrences for 130 nodes; see Supplementary Information and 'Data availability'). Fossil assignment to nodes was based upon synapomorphies for that node from published phylogenetic studies, diagnostic characters for taxonomic ranks and/or detailed surveys of clade fossil records by experts. Fossil ages were used as minimum age constraints; maximum ages were derived for all nodes following the Whole Tree Extension of the Hedman algorithm³⁶, a probabilistic method that incorporates outgroup ages and that has recently been implemented in R³⁷. We identified 130 nodes that could be assigned fossil constraints (Supplementary Information) and time-calibrated the phylogeny using treePL³⁸. A graphical summary of the distribution of calibrations across the phylogeny is shown in Extended Data Fig. 1.

Phylogenetic placement of unsampled species by taxonomy. Using the timecalibrated phylogeny as a backbone, we generated a distribution of trees in which missing taxa were placed according to their taxonomy. For each of the unsampled species of ray-finned fish, we assigned the most restrictive taxonomic rank (for example, genus, family, order) that was recovered as monophyletic in our maximum likelihood phylogeny. To determine divergence times for unsampled species in the phylogeny, we sampled from a distribution of waiting times conditioned on rank-specific estimates of the speciation rate and sampling fraction using a custom Python script implementing functions from TreePar and SimTree $^{39-41}$, and added unsampled species based on the assigned taxonomic rank and inferred waiting time. This procedure was repeated 100 times to generate a distribution of fully sampled ray-finned fish phylogenies (Supplementary Information). This procedure is similar to stochastic polytomy resolution as implemented in PASTIS⁴², but permits construction of extremely large phylogenies using all molecular data in a single analysis, rather than a two-stage process that begins with a reduced backbone dataset followed by separate tree searches for each crown lineage. Additionally, our procedure generates a local estimate of diversification rates at each taxonomic node, rather than using a global diversification rate, permitting more accurate placements of unsampled taxa when diversification rate heterogeneity may exist. Estimation of geographic ranges and species richness. We used the AquaMaps algorithm 43,44 to estimate geographic ranges for marine fishes. These maps were generated using an environmental envelope approach that predicts species distributions based on available species-specific occurrence records at the 0.5°-gridcell scale in conjunction with the following environmental predictors: depth, sea surface temperature, salinity, proportional ice cover and primary productivity⁴⁵. The predictive algorithm also incorporated geographical bounding boxes to limit occurrences to known ocean-scale distributions for each taxon. We transformed the AquaMaps distributions to a Behrmann equal area projection, and upscaled the resulting grids to 150 \times 150-km resolution. We then converted the AquaMaps suitability scores for each cell to binary presence or absence by applying a fixed threshold of 0.5. This threshold was selected based on manual inspection of a number of individual species ranges. Expert opinion was then used when available to further refine the projected distributions, typically by truncating the AquaMaps predictions in light of museum occurrence data, known biogeographical barriers and specialist literature on particular taxa. Where available, we incorporated more accurate distributional maps produced by taxonomic experts in particular groups^{46–48}. The final dataset included maps for 12,050 marine species out of an estimated total of 15,500 described marine species⁴⁹. Our conclusions should be unaffected by these missing and uncommon taxa, given that we were able to reconstruct the previously hypothesized pattern of marine fish richness on a global scale^{3,4}.

Occurrence-based analyses. As an alternative to range predictions from AquaMaps suitability scores, we performed a parallel set of occurrence-based analyses in which we reconstructed cell-based species assemblages as well as species latitudinal midpoints. We obtained all actinopterygian records from four major biodiversity occurrence aggregators (Global Biodiversity Information Facility (GBIF), Ocean Biogeographic Information System (OBIS), Fishnet2 and VertNet) between February 2014 and January 2015 and removed redundancies, resulting in a total of 13,322,575 marine fish occurrences. We downloaded all actinopterygian data from GBIF (https://www.gbif.org/) using their download API version 1; FishNet2 data (http://www.fishnet2.net/) were acquired using a custom Python script to download KML files for each species. VertNet (http://www.vertnet.org/) and OBIS (http://www.iobis.org/) data were retrieved by contacting the administrators of these databases, who then provided us with the relevant data. To reconcile and combine the four datasets, we used museum accession numbers to deduplicate identical records contained in multiple databases. Where accession identifiers were inconsistent within a single museum, we unified these accessions onto a common scheme using a custom Python script. To reconcile species names by resolving synonyms and other sources of error, we used the same procedure described in the Supplementary Information. Institutions contributing substantially to the occurrence dataset are listed in Supplementary Table 6.

The occurrence dataset was filtered to exclude records that fell on land, and records with zero–zero or other nonsensical coordinates. Species richness counts were then calculated across a global grid at 300 \times 300 km resolution, using the Behrmann equal area projection. We further excluded isolated grid cells with recorded species, and removed cells that were greater than two standard deviations from the residuals of a thin plate spline interpolation that was fit to the species richness grid. These filters allowed us to remove cells that were probably unrealistic representations of the species diversity at those locations. For all analyses presented, the same richness and bathymetry filters were applied that were used with the primary map data.

Estimates of speciation rate. We reconstructed speciation rates using (i) an inverse equal-splits measure of speciation rate ($\lambda_{\rm DR}$), also known as the 'DR statistic' ^{21,50,51}, (ii) BAMM estimates of speciation rate allowing time-varying rate regimes ($\lambda_{\rm BAMM}$)^{20,52,53}, and (iii) BAMM estimates of speciation rate assuming time constancy of speciation rates within rate regimes ($\lambda_{\rm BAMM-TC}$). For the $\lambda_{\rm DR}$ analyses, we accounted for missing taxa by computing $\lambda_{\rm DR}$ for each tip in the ATA 31,526 taxon phylogenies; we then computed the average value for each taxon across the set of 100 trees generated with stochastic polytomy resolution. Stochastic polytomy resolution generates taxonomic placements that may compromise inferences of trait-dependent diversification because taxa are placed on trees in a manner that is inconsistent with the underlying process of trait evolution⁵⁴, and we excluded all taxa lacking genetic data from formal statistical analysis of the relationship between latitude and speciation. However, including these taxa during estimates of $\lambda_{\rm DR}$ reduces bias due to incomplete taxon sampling and our calculations of $\lambda_{\rm DR}$ effectively integrate over the number and location of unsampled species.

BAMM analyses were performed on the time-calibrated phylogeny containing 11,638 tips for which genetic data were available. For each of the two classes of BAMM models ($\lambda_{\rm BAMM}$ and $\lambda_{\rm BAMM-TC}$), we performed three BAMM runs for 50 million generations using default MCMC operators and a prior expectation of 500 shifts to facilitate convergence⁵⁵. Raw output and control files to repeat these analyses are available through the Dryad data repository (see 'Data availability'). We were unable to achieve satisfactory convergence when running BAMM on the all-taxon (31,526 tip) phylogenies; we therefore used sampling fractions to account for the effects of incomplete sampling. We corrected for incomplete sampling at the family level. We computed the mean of the marginal posterior distribution of speciation rates for each tip in the phylogeny for both λ_{BAMM} and $\lambda_{BAMM-TC}$. As an alternative to $\lambda_{\rm BAMM}$ and $\lambda_{\rm DR}$, we computed a simple node density estimate of speciation rate²². For each taxon, these estimates are computed simply as n_T/T , in which n_T is the number of nodes on a path of length T, traversing the tree backwards from the tips towards the root. An estimate for an interval of 5 Myr would represent the average speciation rate for a given tip during the past 5 Myr. We computed node density estimates of speciation rate for a sequence of intervals between 0.25 and 50 Myr (Extended Data Fig. 6 and Supplementary Information). As for $\lambda_{\rm DR}$, the node density estimates of speciation rate were computed over the full set of ATA phylogenies.

Grid-based analyses of speciation rate. We computed mean speciation rates ($\lambda_{\rm DR}, \lambda_{\rm BAMM}$ and $\lambda_{\rm BAMM-TC}$) for regional assemblages of fishes, focusing on sets of species that are presumed to occur together at the scale of the 150 \times 150-km grid cell. We computed the mean rate for individual grid cells four different ways, to reduce spatial and taxonomic pseudoreplication across cells. The simplest approach involved computing the arithmetic mean λ for all species inferred to be present in a particular cell (Fig. 1). Following Jetz et al. ²¹, we computed weighted arithmetic and geometric means of speciation rates to reduce the contribution of geographically widespread taxa to the overall mean. For the arithmetic mean, the rate for the kth

grid cell is computed as $\lambda_k = (\Sigma w_i \lambda_i)/\Sigma w_i$ in which Σ denotes a summation over all N species (i=1 to i=N) present in the kth cell, and w_i is the weight assigned to the ith species. We computed weights for each species as the inverse of the number of grid cells in which the species was found²¹. Therefore, geographically widespread taxa contribute less to a cell mean than a taxon with a highly restricted geographical distribution. Finally, we computed cell means for 'realm endemics'—species uniquely found in one of 12 biogeographical realms under the MEOW marine bioregionalization scheme⁵⁶ (n=1,053 and 3,100 realm endemics with and without genetic data). The analysis of endemic taxa is particularly informative as such taxa provide more localized information about speciation rates in particular geographical regions relative to widespread taxa that may be found in multiple regions⁵⁷.

To formally assess the relationship between latitude and assemblage-level speciation rate, we first computed the mean speciation rate for all cells within a particular ecoregion (n = 262) under the MEOW biogeographical regionalization⁵⁶. We modelled the speciation-latitude relationship at the scale of ecoregions rather than individual cells because of the high autocorrelation between adjacent cells, which was reduced at the ecoregion scale, and to reduce the computational burden associated with analysing the full (16,150 grid cell) dataset. To account for spatial autocorrelation between ecoregions, we implemented simultaneous autoregressive error (SAR) models using the spdep package in R⁵⁸⁻⁶⁰. These and other statistical tests are two-tailed. We defined neighbours for SAR models as those ecoregions with contiguous boundaries; we then selected the appropriate weighting scheme using Akaike information criterion (AIC) model selection. Simple visual inspection of our data (Fig. 1 and Extended Data Fig. 2) and ordinary least squares (OLS) breakpoint regressions reveal a clear biphasic signal in the relationship between speciation rate and latitude, with a linear increase for higher latitude cells (approximately 30° N and 30° S) and a much weaker relationship for low (tropical) latitudes. We therefore considered an expanded set of breakpoint SAR models with no relationship between absolute latitude and speciation for cells below a particular threshold value, and a linear effect of absolute latitude on speciation above the threshold. We used maximum likelihood analyses to estimate the threshold location and we compared the fit of the breakpoint model to a simple no-breakpoint SAR model using AIC (Extended Data Fig. 2). We used Moran's I to test for spatial autocorrelation in the residuals of OLS and SAR regressions to determine whether the SAR model successfully accounted for spatial non-independence in the data. We tested the relationship between assemblage speciation rate and latitude for ecoregions with absolute latitude less than the previously identified breakpoints (for example, tropical and other low-latitude regions). In general, there is at most a marginal effect of latitude on speciation rate for tropical and subtropical regions (Extended Data Fig. 2h). Finally, we estimated endemism for each MEOW marine biogeographical province using two analyses of occupancy. These two analyses of regional endemism, E, are given by $E = (1/N)\Sigma(1/O_k)$, in which N is the number of species occurring in the focal region, O_k is the estimated global occupancy of the *k*th species from that region, and Σ denotes a summation from k=1 to k=N. Occupancy is computed as either the total number of biogeographical provinces or as the total number of equal-area grid cells in which the taxon is found.

Trait-dependent speciation. We treated the absolute value of the latitudinal midpoint of each species as a 'trait' and tested its relationship to speciation rates using formal statistical methods for analysing trait-dependent diversification 26,51. The latitudinal midpoint for each species was computed as the centroid midpoint of the geographical range of the species. We used three recently developed methods for testing the effects of species traits on lineage speciation rates that are robust to $phylogenetic\ pseudoreplication\ and\ model\ misspecification^{61,62}.\ Using\ ES-SIM^{25},$ we tested whether λ_{DR} was correlated with absolute latitudinal midpoint for individual species. Using FiSSE⁵¹, we then tested whether two discrete classifications of species by latitude ('low latitude/tropical' versus 'high latitude/temperate') differed in their rate of speciation as measured using λ_{DR} . We performed the FiSSE test across a range of thresholds for classifying lineages as tropical and temperate (23.5°, 25°, 30°, 35°, 40°, 45°, 50°, 55° and 60°). Regardless of the threshold, all FiSSE results indicated a highly significant effect of latitude on speciation rate (Extended Data Table 1). As an alternative method for continuous traits, we used STRAPP26 to test whether latitude was correlated with the two BAMMbased measures of speciation rate ($\lambda_{\rm BAMM}$ and $\lambda_{\rm BAMM-TC}$). Results for $\lambda_{\rm BAMM}$ and $\lambda_{BAMM-TC}$ were almost identical and identified a strong effect of latitude on speciation rate (Pearson's r = 0.30-0.31; $P \le 0.006$). One possible explanation for our results is that high-latitude assemblages are enriched for deep-sea taxa, and that faster speciation is actually a property of deep-sea environments and not high latitudes. To test this hypothesis, we obtained depth classifications for marine fishes from FishBase (http://www.fishbase.org); minimum and maximum depths were available for 4,089 species (of 5,231 total species). We used ES-SIM to test the relationship between latitude and speciation $\lambda_{\rm DR}$ for fishes with minimum depth >200 m. Using FiSSE, we tested whether speciation rates were faster for low-latitude deep-sea fishes relative to low-latitude shallow or reef-associated species (Extended Data Fig. 7a).

Additional checks on statistical robustness. We performed several additional checks on the robustness of the latitude–speciation correlation. We visualized latitudinal trends in terminal branch length, which is expected to correlate inversely with underlying speciation rate. We obtained estimates of the mean terminal branch length of each species from the distribution of ATA (31,526 taxon) phylogenies. The inverse of these branch lengths is the simplest possible estimate of the instantaneous rate of speciation, although it is an extremely noisy metric; $\lambda_{\rm DR}$ is similar but includes the weighted contribution of earlier branches to increase the signal-to-noise ratio. Despite the overall noisiness of the terminal branch length metric, there is a clear trend towards shorter terminal branches for high-latitude taxa (Extended Data Fig. 9).

We also tested whether our speciation rate estimates could have been driven by latitudinal gradients in genetic taxon sampling, as might be the case if a higher percentage of high-latitude taxa had DNA sequences with which to infer their phylogenetic position without relying on stochastic polytomy resolution. To formally address this potential confounding variable, we fitted multiple regression models to the relationship between lineage-specific speciation rate and latitudinal midpoint, but including the sampling fraction for each species as a covariate. The sampling fraction for each taxon was simply the percentage of total species from the corresponding family-level clade that contained genetic data (for example, the percentage of total species from each family that were represented in the genetic supermatrix). These sampling fractions were identical to those used to correct for incomplete sampling in the BAMM analyses. Visual inspection and formal analysis shows minimal effect of sampling fraction on the patterns reported here (Extended Data Fig. 9).

Finally, we tested whether variation in the rate of molecular evolution could drive spurious variation in the rate of diversification. A systematic bias towards low rates of molecular evolution can lead to apparent fast rates of diversification in slowly evolving lineages as an artefact of the algorithms used for time-scaling the raw (uncalibrated) phylogeny. If our results are affected by this bias, we expect to observe (i) a general trend towards slower rates of molecular evolution at high latitudes, and (ii) a negative correlation between speciation rate and the rate of molecular evolution. To estimate rates of molecular evolution, we computed the relative root-to-tip distances for each taxon in the phylogeny and estimated their correlation with both latitude and $\lambda_{\rm DR}$ (Extended Data Fig. 9). There is no evidence that higher latitudes are associated with slower rates of molecular evolution, or that rates of molecular evolution are negatively correlated with $\lambda_{\rm DR}$ (see Extended Data Fig. 9).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. All scripts and code necessary to repeat the analyses described here have been made available through the Dryad digital data repository (https://doi.org/10.5061/dryad.fc71cp4).

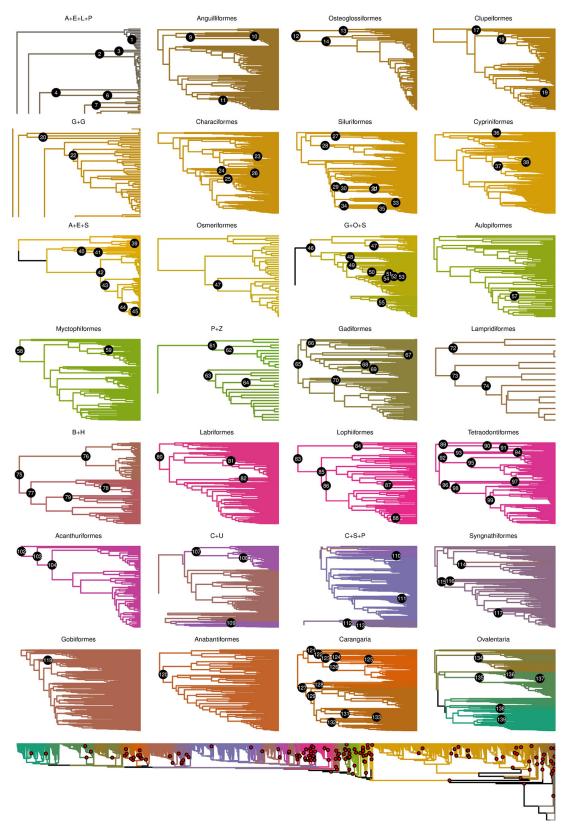
Data availability. All data necessary to repeat the analyses described here have been made available through the Dryad digital data repository (https://doi.org/10.5061/dryad.fc71cp4). Phylogenetic tree distributions are also available through http://fishtreeoflife.org.

- Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. 9, 37 (2009).
- 31. Betancur-R, R. et al. The tree of life and a new classification of bony fishes. *PLoS Curr. Tree of Life* https://doi.org/10.1371/currents. tol.53ba26640df0ccaee75bb165c8c26288 (2013).
- Near, T. J. et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl Acad. Sci. USA 110, 12738–12743 (2013).
- Aberer, A. J., Krompass, D. & Stamatakis, A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Syst. Biol. 62, 162–166 (2013).
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder
 new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773 (2017).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).
- Hedman, M. M. Constraints on clade ages from fossil outgroups. *Paleobiology* 36, 16–31 (2010).
- Lloyd, G. T., Bapst, D. W., Friedman, M. & Davis, K. E. Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biol. Lett.* 12, 20160609 (2016).
- Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690 (2012).
- Stadler, T. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theor. Biol. 261, 58–66 (2009).
- Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl Acad. Sci. USA 108, 6187–6192 (2011).
- Stadler, T. Simulating trees with a fixed number of extant species. Syst. Biol. 60, 676–684 (2011).



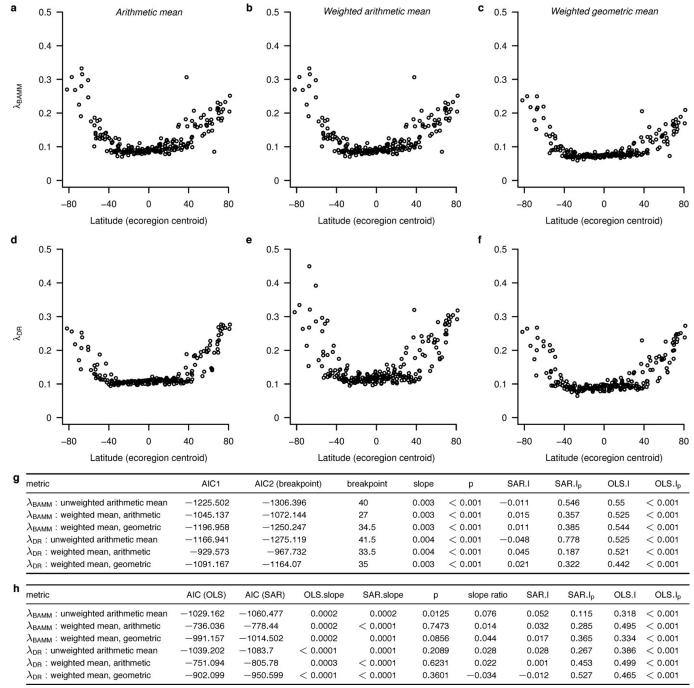
- 42. Thomas, G. H. et al. PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods Ecol. Evol.* **4**, 1011–1017 (2013).
- 43. Ready, J. et al. Predicting the distribution of marine organisms at the global scale. *Ecol. Modell.* **221**, 467–478 (2010).
- 44. Kaschner, K. et al. AquaMaps: predicted range maps for aquatic species. version 08/2016 http://www.aquamaps.org (2016).
- 45. Kaschner, K. et al. AquaMaps environmental dataset: half-degree cells authority file (HCAF). version 6, 08/2016 https://www.aquamaps.org/main/envt_data. php (2016).
- Mecklenburg, C. W., Mecklenburg, T. A., Sheiko, B. A. & Steinke, D. Pacific Arctic Marine Fishes (CAFF, Akureyri, 2016).
- Coll, M. et al. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS ONE 5, e11842 (2010).
- IUCN. The IUCN Red List of Threatened Species, version 2016-1 http://www. iucnredlist.org, downloaded on 8 February 2018 (2016).
- Mora, C., Tittensor, D. P. & Myers, R. A. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proc. R. Soc. B* 275, 149–155 (2008).
- Belmaker, J. & Jetz, W. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecol. Lett.* 18, 563–571 (2015).
- Rabosky, D. L. & Goldberg, E. É. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* 71, 1432–1442 (2017).
- 52. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543 (2014).

- Rabosky, D. L., Donnellan, S. C., Grundler, M. & Lovette, I. J. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. Syst. Biol. 63, 610–627 (2014).
- Rabosky, D. L. No substitute for real data: a cautionary note on the use of phylogenies from birth-death polytomy resolvers for downstream comparative analyses. *Evolution* 69, 3207–3216 (2015).
- Mitchell, J. S. & Rabosky, D. L. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods Ecol. Evol.* 8, 37–46 (2017).
- 56. Spalding, M. D. et al. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
- Cowman, P. F., Parravicini, V., Kulbicki, M. & Floeter, S. R. The biogeography of tropical reef fishes: endemism and provinciality through time. *Biol. Rev. Camb. Philos. Soc.* 92, 2112–2130 (2017).
- 58. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. R News 5, 9–13 (2005).
- Bivand, R., Hauke, J. & Kossowski, T. Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geogr. Anal.* 45, 150–179 (2013).
- Bivand, R. & Piras, G. Comparing implementations of estimation methods for spatial econometrics. J. Stat. Softw. 63, 1–36 (2015).
- Rabosky, D. L. & Goldberg, E. E. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64, 340–355 (2015).
- Maddison, W. P. & FitzJohn, R. G. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64, 127–136 (2015).



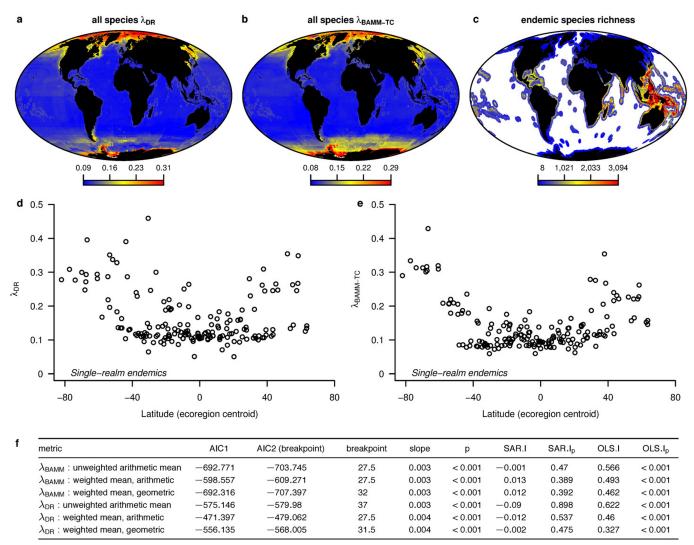
Extended Data Fig. 1 | Phylogenetic placement of fossil calibrations in major fish lineages. Major lineages are broken into subclades (top) to visualize fossil calibrations and are coloured by taxonomic order. Numbered nodes are described in the calibration report in the Dryad data repository. The same calibrations are red circles in the full phylogeny (bottom). A + E + L + P: Acipenseriformes, Elopiformes,

Lepisosteiformes, Polypteriformes; A+E+S: Argentiniformes, Esociformes, Salmoniformes; B+H: Beryciformes, Holocentriformes; C+S+P: Centrarchiformes, Scombriformes, Perciformes; C+U: Chaetodontiformes, Uranoscopiformes; C+G: Gonorynchiformes, Gymnotiformes; C+G: Galaxiiformes, Osmeriformes, Stomiatiformes; C+G: Percopsiformes, Zeiformes.



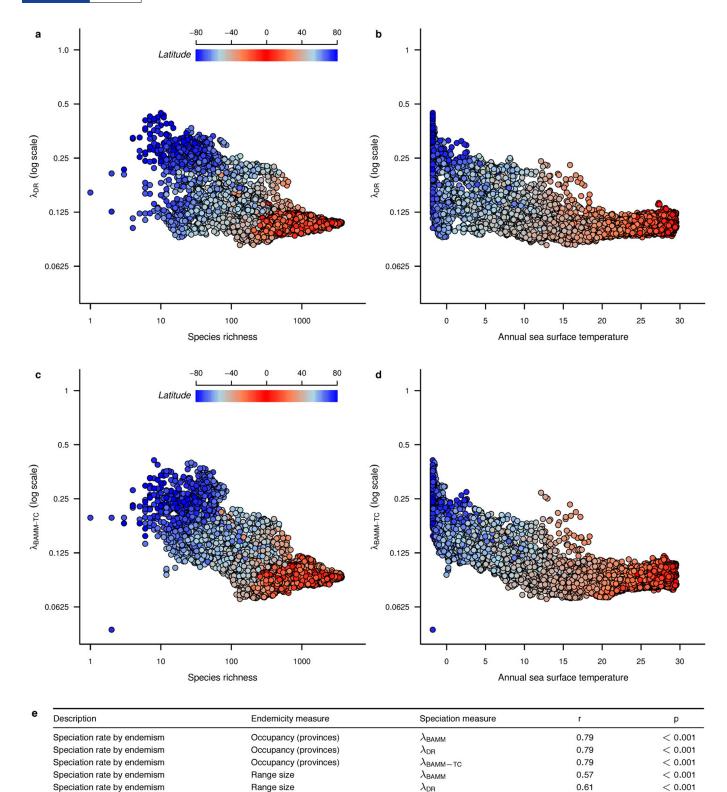
Extended Data Fig. 2 | Relationships between mean speciation rates and latitude for 262 marine ecoregions using alternative methods for the computation of the cell rates. a–c, $\lambda_{\rm BAMM}$ versus latitude. d–f, $\lambda_{\rm DR}$ versus latitude. Ecoregion rates are mean rates across all cells assigned to each biogeographical region. Arithmetic mean is the mean rate across all taxa inferred to occur in the cell; weighted arithmetic and weighted geometric means assign proportionately greater weight to species with small geographical ranges. Weighting schemes for speciation metrics are described in the Methods. g, Simultaneous autoregressive (SAR) spatial error models for the effects of absolute latitude on mean speciation rates for ecoregions. AIC1 gives the Akaike information criterion (AIC) for a linear model with a single slope and intercept term; AIC2 is the corresponding AIC for a breakpoint model that assumes no relationship (slope = 0) between absolute latitude and speciation

rate for all values below some threshold, and a linear relationship for latitudes that exceed the threshold. SAR.I and SAR. I_P are global Moran's I estimates and associated P values for assessing the presence of residual spatial autocorrelation in the model residuals; OLS.I and OLS. I_P are the corresponding values for ordinary least squares (OLS) regression that ignores spatial autocorrelation. All SAR models show highly significant effects of latitude on speciation rate, and breakpoint models provided a consistently better fit than models without a breakpoint. h, OLS and SAR models for the effects of absolute latitude on speciation rate for low-latitude grid ecoregions only. The slope ratio term gives the ratio of slopes for low-latitude ecoregions (below the corresponding breakpoint; g) to the slope for ecoregions with latitude above the breakpoint. Overall, there is a marginal effect of latitude on speciation rate for low-latitude ecoregions.



Extended Data Fig. 3 | Relationships between speciation rate and latitude for alternative speciation rate metrics and for endemic taxa only. a, b, Global maps of $\lambda_{\rm DR}$ and $\lambda_{\rm BAMM-TC}$, as in Fig. 1. c, Global map of endemic species richness, by grid cell. 'Endemic' taxa are those that are restricted to a single MEOW realm; an endemic taxon can occur in multiple grid cells provided all grid cells are contained within a single realm. d, Relationship between speciation rates ($\lambda_{\rm DR}$) and latitude for ecoregions (n=232), computed using realm endemics only.

e, Relationship between speciation rates ($\lambda_{\rm BAMM-TC}$) and latitude for ecoregions, computed using realm endemics only. f, SAR spatial error models for the relationship between ecoregion speciation rates and absolute latitude, for which ecoregion means are computed from single-realm endemics only. Weighting schemes for assemblages are described in the Methods. SAR modelling results are presented as in Extended Data Fig. 2g and show a strong correlation between latitude and speciation rate.



Extended Data Fig. 4 Speciation rate, species richness, temperature
and endemism. a, Negative relationship between species richness
and mean speciation rate (λ_{DR}) for individual grid cells. b , Negative
relationship between mean annual sea surface temperature and mean
speciation rate. c , d , Same as a , b , but for BAMM with time-constant
rate regimes ($\lambda_{\text{BAMM-TC}}$). Grid cells as in Fig. 1 ($n = 16,150$). See Fig. 2
for comparison. e , Correlation between mean speciation rate for
MEOW biogeographical provinces and two measurements of regional

Range size

Range size

Occupancy (provinces)

Speciation rate by endemism

Endemism by absolute latitude

Endemism by absolute latitude

endemism. 'Occupancy (provinces)' measures endemism as the inverse of the mean number of provinces occupied by each taxon that occurs in a particular province. 'Range size' is the inverse mean range size across all taxa occurring in a given province. High values of endemism indicate that a given region consists of species that are found in fewer additional provinces, or of species with smaller geographical ranges. The bottom two rows show the correlations between the endemism parameters and latitude.

 $\lambda_{\mathsf{BAMM-TC}}$

0.56

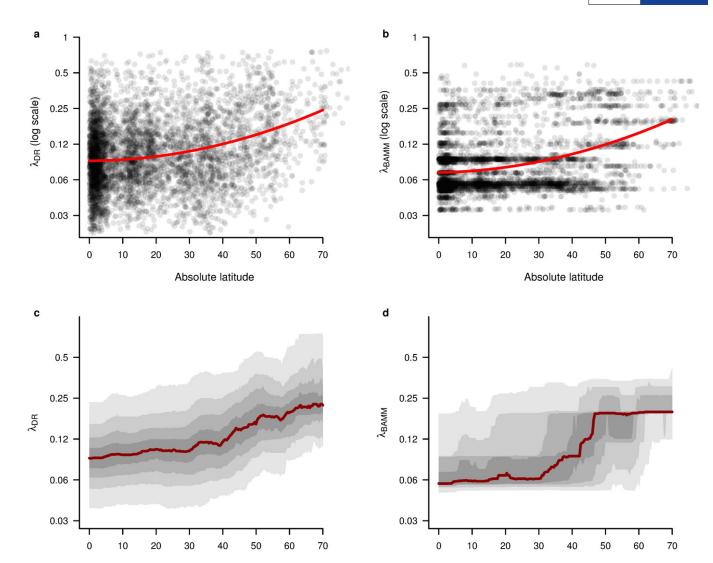
0.58

0.46

< 0.001

< 0.001

< 0.001

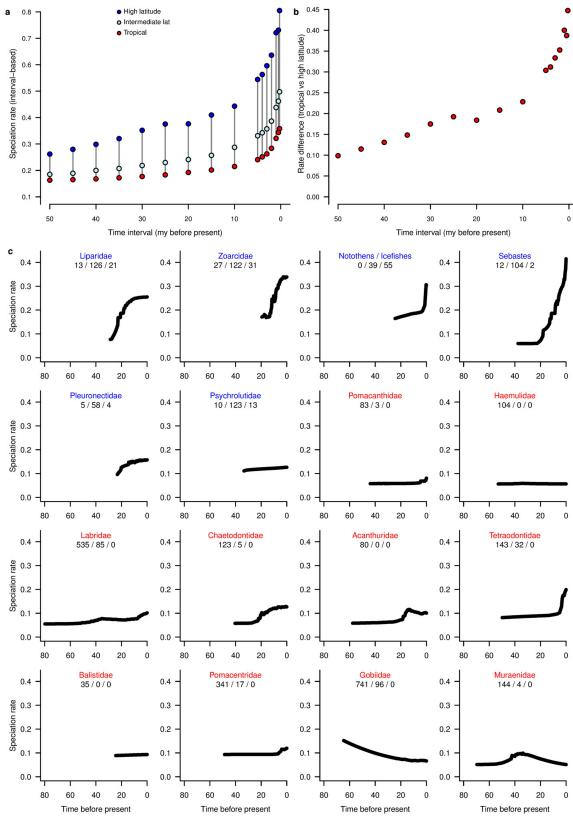


Extended Data Fig. 5 | Speciation rates for individual taxa as a function of latitudinal midpoint. a, $\lambda_{\rm DR}$ for all marine species with genetic data (n=5,229) as a function of the latitudinal (centroid) midpoint of their geographical range. Non-phylogenetic OLS regression with quadratic term is overlaid on points to denote trend in mean rates. b, $\lambda_{\rm BAMM}$ for the same taxon set. c, Sliding window analysis of $\lambda_{\rm DR}$ distributional quantiles

Absolute latitudinal midpoint

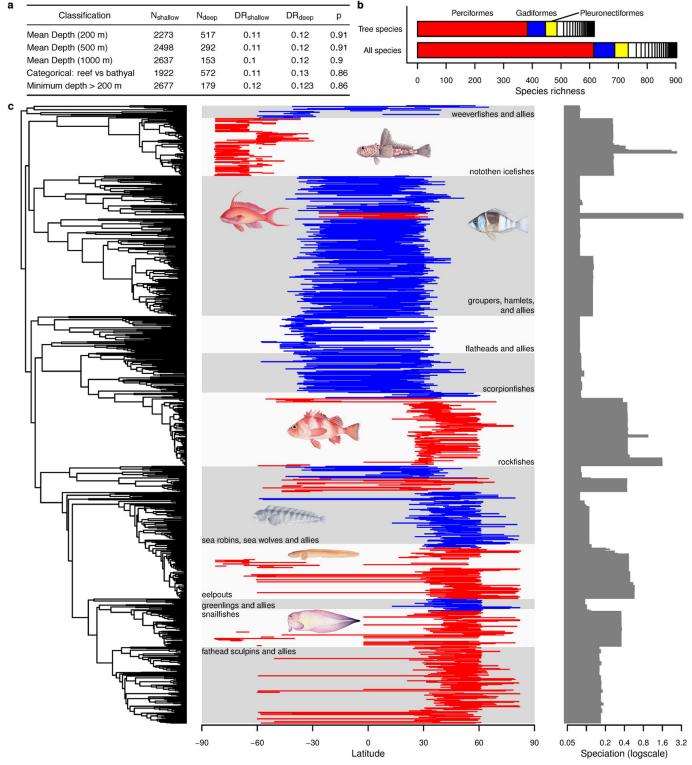
in speciation rates by individual taxa with respect to latitudinal midpoint. Contours denote quantiles from 0.10 to 0.90, in 0.10 increments, with a sliding window size of 6°. Dark red line is the median rate. **d**, Distributional quantiles of $\lambda_{\rm BAMM}$ for all species with respect to latitudinal midpoint; dark red line is median rate.

Absolute latitudinal midpoint



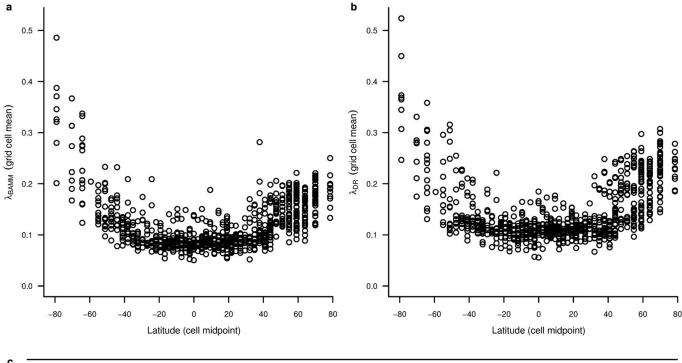
Extended Data Fig. 6 | Temporal dimension of speciation rate variation as a function of latitude. a, Mean speciation rates for taxa from low latitudes ($<30^\circ$), intermediate latitudes ($30-60^\circ$) and high latitudes ($>60^\circ$) computed using the interval method. Per-taxon interval-based rates were computed for time intervals between 0.25 and 50 million years before present. Time-averaged speciation rates for high-latitude fishes are much higher than those inferred for low-latitude fishes, even across timescales that exceed 20 million years. b, Rate differential between high-latitude and low-latitude taxa as a function of interval duration. c, Speciation-

over-time curves reconstructed using the time-varying rates model in BAMM for 14 clades of high-latitude (blue) and low-latitude (red) fishes. Inset numbers for each panel give the numbers of low-, intermediate- and high-latitude (from left to right) taxa from each clade for which geographical range data are available. Low-latitude clades were selected to represent high-diversity and iconic reef-associated clades that contribute substantially to the tropical diversity peak in marine fishes. With the possible exception of gobies, there is no signal of early, rapid speciation in low-latitude or tropical shallow-water clades.



Extended Data Figure 7 | Speciation rates in deep-sea fishes and the phylogenetic structure of high-latitude fish diversity. a, Formal test of the relationship between speciation rate and depth classification for tropical fishes. 'Classification' is the criterion used to define fishes as deep sea versus shallow water; mean depth (200 m) thus classifies all fishes with mean depth greater than 200 m as deep sea. Among tropical fishes, there is no effect of depth state on speciation rate. b, Phylogenetic composition of high-latitude fish diversity by taxonomic order, across all marine fishes (top) and for the subset of species with genetic data (bottom). High latitude is defined as having a centroid midpoint greater than 45° north or south. Only the three most species-rich high-latitude orders are labelled. Most high-latitude marine fishes are Perciformes. c, Phylogenetic and geographical structure of the diversity of Perciformes.

The latitudinal range of each perciform species in the phylogenetic dataset is shown, along with the corresponding speciation rate ($\lambda_{\rm BAMM}$). Latitudinal ranges from species with speciation rates that are faster and slower than the median rate are shown in red and blue, respectively. High-latitude and rapidly speciating clades are nested within slowly speciating tropical lineages, and speciation rates for high-latitude taxa of Perciformes are higher than those observed in tropical lineages. Mean speciation rates for high-latitude species (>45°, n=376) are faster than those observed for tropical (<25°, n=287) species (tropical: $\lambda_{\rm DR}=0.16,\,\lambda_{\rm BAMM}=0.15;$ high latitude: $\lambda_{\rm DR}=0.30,\,\lambda_{\rm BAMM}=0.23$). For polar species (>60°, n=105), these rate differentials are even more extreme, with mean $\lambda_{\rm DR}=0.38$ and $\lambda_{\rm BAMM}=0.31$.

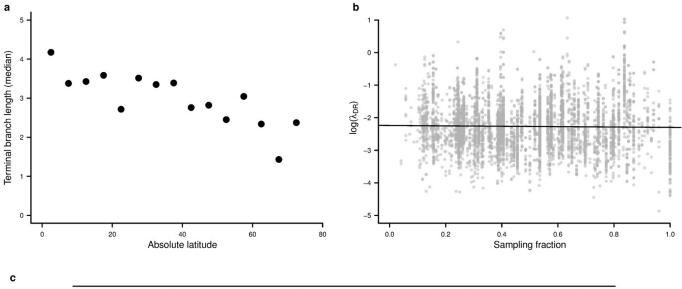


С	metric	AIC1	AIC2 (breakpoint)	breakpoint	slope	р	SAR.I	SAR.Ip	OLS.I	OLS.Ip
	λ_{BAMM}	-861.634	-892.706	32.5	0.0031	0	-0.0143	0.5550	0.5094	1.37E-16
	λ_{DR}	-805.025	-848.045	34	0.0038	0	-0.0308	0.6545	0.5623	2.76E-19

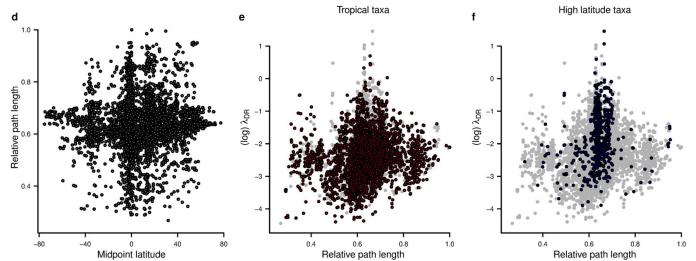
d	Threshold	λ_1	λ_{0}	р	Null $\Delta\mu$	Null σ	n _{parsimony}	q _{parsimony}
	30	0.1194	0.2057	< 0.001	0.0001	0.0254	398	0.0044
	35	0.1232	0.2194	< 0.001	0.0009	0.0276	303	0.0031
	40	0.1270	0.2306	< 0.001	0.0007	0.0310	218	0.0021
	45	0.1285	0.2528	< 0.001	-0.0004	0.0316	178	0.0017
	50	0.1310	0.2682	< 0.001	0.0004	0.0323	171	0.0017
	55	0.1352	0.3019	< 0.001	0.0006	0.0356	108	0.0010
	60	0.1370	0.3789	< 0.001	-0.0010	0.0419	65	0.0005

Extended Data Figure 8 | Latitudinal gradient in speciation rate for cell assemblages inferred from occurrence data. Cell assemblages (n=843) and species latitudinal midpoints were inferred from a non-redundant merge of four primary occurrence-based biodiversity databases (GBIF, OBIS, Fishnet2 and VertNet). **a**, $\lambda_{\rm BAMM}$ for cell assemblages as a function of latitude. **b**, $\lambda_{\rm DR}$ as a function of latitude. **c**, SAR spatial error models for the effects of absolute latitude on mean speciation rates for grid cells. AIC1 is a linear model with a single slope and intercept term; AIC2 is the corresponding AIC for a breakpoint model that assumes no relationship (slope = 0) between absolute latitude and speciation rate for all values below some threshold, and a linear relationship for latitudes that exceed the threshold. All other column headings as in Extended

Data Fig. 2g. Results indicate a strong effect of latitude on speciation rate and are nearly identical to results obtained using the dataset of the primary map. **d**, Effects of absolute latitudinal midpoint for individual taxa on corresponding tip speciation rates, as assessed using FiSSE. Each row gives the results of FiSSE using a different threshold for classifying lineages as tropical and temperate. λ_0 and λ_1 denote estimated speciation rates (similar to $\lambda_{\rm DR}$) for tropical and temperate lineages, respectively. All column headings are identical to those shown in Extended Data Table 1. Results are nearly identical to those obtained using explicit range reconstructions and reveal a pervasive effect of latitude on lineage-level speciation rates, regardless of the threshold used to classify species.



· ·					77
		Variable	Slope	Sum of Squares	P-value
	λ_{BAMM}	latitude	0.012	255.96	< 0.0001
		sampling frac.	0.075	1.45	0.048
	λ_{DR}	latitude	0.011	206.71	< 0.0001
		sampling frac.	-0.103	2.74	0.021



Extended Data Fig. 9 | Additional checks of statistical robustness. a, Relationship between terminal branch lengths and absolute latitudinal midpoint; means are shown for all species falling into a given bin $(\pm 2.5^{\circ})$ from the focal value, n = 15). Mean branch lengths decrease with increasing latitude, reflecting faster speciation at high latitudes. **b**, Relationship between the estimated speciation rate for each taxon $(\lambda_{\rm DR}, n = 5,155)$ and the sampling fraction for the corresponding familylevel clade to which the taxon belongs; the sampling fraction is simply the percentage of known taxa from the family that were represented in the phylogenetic dataset with genetic data. There is no clear relationship between the sampling fraction and the estimated speciation rates. c, Multiple regression analysis (OLS) of the relationship between taxonspecific speciation rate ($\lambda_{\rm BAMM}$ or $\lambda_{\rm DR}$) and two predictors (latitude and family-level sampling fraction) in a multiple regression framework (n = 5,155). If the relationship between speciation rate and latitude is driven by progressively greater (or lower) genetic taxon sampling as a function of latitude, the sampling fraction term should explain a large

fraction of the overall sums of squares. Even when sampling fraction is included as a covariate, the overwhelming fraction of variance is explained by latitude. For both λ_{DR} and λ_{BAMM} , more than 98% of the total sums of squares is explained by latitude and not sampling. $\mathbf{d} - \mathbf{f}$, Test for the effects of molecular evolutionary rate variation and latitudinal bias in speciation rate. d, Relationship between root-to-tip branch length sum for uncalibrated (non-ultrametric) RAxML phylogeny and midpoint latitude for each marine taxon (n = 5,149). **e**, **f**, Relationship between root-to-tip distance and λ_{DR} . There is effectively no relationship between the total path length for individual tips and their absolute latitudinal midpoint (Pearson r = 0.020). Plots in **e** and **f** emphasize tropical (midpoint latitude $<25^{\circ}$; n = 3,481; red) and temperate-polar (midpoint latitude $>45^{\circ}$; n = 567; blue) taxa, respectively, all other taxa are shown in grey. Overall relationship between (log) λ_{DR} and the rate of molecular evolution (rootto-tip sum) is weak but positive (Pearson r = 0.130) and inconsistent with the hypothesis that slow rates of molecular evolution at high latitudes results in fast but spurious estimates of speciation rate.



Extended Data Table 1 | Effects of absolute latitudinal midpoint on speciation rates

Threshold	λ_1	λ_{0}	р	Null $\Delta\mu$	Null σ	n _{parsimony}	q _{parsimony}
23.5	0.119	0.187	< 0.001	0.00122	0.022	600	0.007
25	0.12	0.19	< 0.001	-0.00098	0.022	576	0.007
30	0.122	0.2	< 0.001	-0.00035	0.023	468	0.005
35	0.126	0.205	< 0.001	-0.00007	0.025	388	0.004
40	0.128	0.223	< 0.001	0.00043	0.028	263	0.002
45	0.129	0.246	< 0.001	0.00039	0.03	201	0.002
50	0.132	0.267	< 0.001	0.00034	0.031	185	0.002
55	0.136	0.275	< 0.001	0.00174	0.034	132	0.001
60	0.137	0.353	< 0.001	0.00061	0.041	74	0.001

The effect of latitude on diversification was assessed using FiSSE, a method for inferring the effects of a binary character on lineage diversification rates. Each row gives the results of FiSSE using a different threshold for classifying lineages as tropical and temperate. λ_0 and λ_1 denote estimated speciation rates (similar to λ_{DR}) for tropical and temperate lineages, respectively. P values indicate the proportion of simulations with a rate difference ($\lambda_1 - \lambda_0$) that is greater than the observed difference ($\Delta\mu_{null}$). The number of parsimony-reconstructed changes between states 0 and 1 is given by $\eta_{parsimony}$; q denotes the empirically estimated transition rate used to generate the null distribution. Results are based on 2,000 simulations; the observed difference in rates exceeded all simulated values, regardless of threshold.



Hot streaks in artistic, cultural, and scientific careers

Lu Liu^{1,2,3}, Yang Wang^{1,2}, Roberta Sinatra^{4,5,6}, C. Lee Giles^{3,7}, Chaoming Song⁸ & Dashun Wang^{1,2,9}*

The hot streak-loosely defined as 'winning begets more winnings'—highlights a specific period during which an individual's performance is substantially better than his or her typical performance. Although hot streaks have been widely debated in sports^{1,2}, gambling³⁻⁵ and financial markets^{6,7} over the past several decades, little is known about whether they apply to individual careers. Here, building on rich literature on the lifecycle of creativity⁸⁻²², we collected large-scale career histories of individual artists, film directors and scientists, tracing the artworks, films and scientific publications they produced. We find that, across all three domains, hit works within a career show a high degree of temporal regularity, with each career being characterized by bursts of high-impact works occurring in sequence. We demonstrate that these observations can be explained by a simple hot-streak model, allowing us to probe quantitatively the hot streak phenomenon governing individual careers. We find this phenomemon to be remarkably universal across diverse domains: hot streaks are ubiquitous yet usually unique across different careers. The hot streak emerges randomly within an individual's sequence of works, is temporally localized, and is not associated with any detectable change in productivity. We show that, because works produced during hot streaks garner substantially more impact, the uncovered hot streaks fundamentally drive the collective impact of an individual, and ignoring this leads us to systematically overestimate or underestimate the future impact of a career. These results not only deepen our quantitative understanding of patterns that govern individual ingenuity and success, but also may have implications for identifying and nurturing individuals whose work will have lasting impact.

According to the Matthew effect^{9,23,24}, victories bring reputation and recognition that can translate into tangible assets, which in turn help to bring future victories. This school of thought supports the existence of a hot streak in a career, which is also consistent with literature in the field of innovation showing that peak performance clusters in time, typically occurring around the middle of a career^{8,11,21}. On the other hand, the random impact rule uncovered in the arts^{10,21} and sciences^{10,18} predicts the opposite: the best works occur randomly within a career, and their occurrence is primarily driven by productivity. Following this school of thought, works after a major breakthrough are not affected by what preceded them, supporting the viewpoint of regression towards the mean. The two divergent schools of thought raise a fundamental question: do hot streaks exist in creative careers?

To answer this question, we collected data sets recording the career histories of individual artists, film directors and scientists (Supplementary Information S1) and traced the impact of the artworks, films and papers they produced, approximated by auction prices ¹⁵, IMDB ratings (https://www.imdb.com/)²⁵ and citations garnered after 10 years of publication $(C_{10})^{13,16,18,26}$, respectively (see Methods). We started by investigating the timing of the three most impactful works

produced in each career. In a sequence of N works by an individual, we denoted with N^* the position of the highest-impact work within a career, N^{**} the second highest and N^{***} the third. We found that each of the three highest-impact works was randomly distributed among all the works produced by an individual (Extended Data Fig. 1a–c), offering strong endorsement for the random impact rule 10,18,21 .

However, as we show next, the randomness in individual creativity is only apparent, because the timing between creative works follows highly predictable patterns. We measured the correlation between the timing of the two biggest hits within a career, and compared it with a null hypothesis in which N^* and N^{**} each occured at random. The normalized joint probability, $\phi(N^*,N^{**})=P(N^*,N^{**})/(P(N^*)P(N^{**}))$, is substantially overrepresented along the diagonal elements of matrices (Fig. 1a–c), demonstrating that N^* and N^{**} are much more likely to colocate with each other than would be expected from the random impact model across three domains. The diagonal pattern disappears if we shuffle the order of works within each career, thereby breaking the temporal correlations (Extended Data Fig. 1j–r).

To quantify the temporal colocation of hits observed in Fig. 1a–c, we calculated the distance between the two highest-impact works for every individual, measured by the number of works produced in between, $\Delta N = N^* - N^{**}$. We compared $P\left(\frac{\Delta N}{N}\right)$ of real careers with $P_{\rm S}\left(\frac{\Delta N}{N}\right)$ of shuffled careers by defining $R\left(\frac{\Delta N}{N}\right) = P\left(\frac{\Delta N}{N}\right)/P_{\rm S}\left(\frac{\Delta N}{N}\right)$. For artists, directors, and scientists, all $R\left(\frac{\Delta N}{N}\right)$ exhibit a clear peak centring around zero and decay quickly as ΔN deviates from zero (Fig. 1d–f). Notably, $R\left(\frac{\Delta N}{N}\right)$ is mostly symmetric around zero (Fig. 1d–f), indicating that the biggest hit is equally likely to arrive before or after the second biggest. The colocation patterns are not limited to the two highest-impact works within a career. We repeated our analyses for other pairs of hit works, such as N^{**} versus N^{***} and N^{**} versus N^{***} , and uncovered the same colocation patterns (Extended Data Fig. 1d–i).

Do high impact works come in streaks within a career? We counted the number of consecutive works whose impact exceeded the median of all works within a career (Extended Data Fig. 2d–f). We calculated the length of the longest streak L for each career. We then shuffled the order of works within each career, and measured again their longest streaks $L_{\rm s}$. P(L) was characterized by a much longer tail than $P(L_{\rm s})$ (Fig. 1g–i), indicating that real careers are characterized by long streaks of relatively high-impact works clustered together in sequence. We tested the robustness of these results by controlling for individual career length, and by varying our threshold used to calculate L, and arrived at the same conclusions (Extended Data Figs. 2–4, Supplementary Information S2). Together, these results raise an important question: what mechanisms are responsible for the temporal regularities observed across diverse career histories?

Let us first consider a null model in which the goodness of works produced in a career (that is, log(price) for artists, ratings for directors,

¹Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA. ²Kellogg School of Management, Northwestern University, Evanston, IL, USA. ³College of Information Sciences and Technology, Pennsylvania State University, University Park, PA, USA. ⁴Department of Network and Data Science, and Department of Mathematics and its Applications, Central European University, Budapest, Hungary. ⁵Center for Complex Network Research, Northeastern University, Boston, MA, USA. ⁶Complexity Science Hub, Vienna, Austria. ⁷Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA. ⁸Department of Physics, University of Miami, Coral Gables, FL, USA. ⁹McCormick School of Engineering, Northwestern University, Evanston, IL, USA. *e-mail: dashun.wang@northwestern.edu

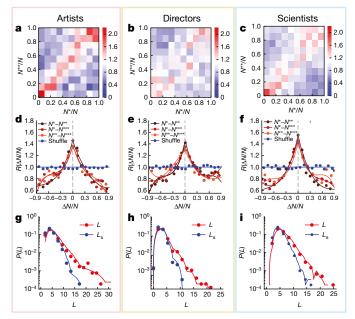


Fig. 1 | Hot streaks in artistic, cultural and scientific careers. \mathbf{a} - \mathbf{c} , $\phi(N^*, N^{**})$, colour coded, measures the joint probability of the two highestimpact works within a career for artists (a), directors (b), and scientists (c). $\phi > 1$ indicates that two hits are more likely to colocate than would be expected at random. \mathbf{d} - \mathbf{f} , $R(\frac{\Delta N}{N})$ measures the temporal distance between highest-impact works relative to the null model's prediction. Real careers show a clear peak around 0 (red dots), which is well captured by the hotstreak model (solid lines). Different shades of red correspond to different pairs of hit works. Blue dots denote the same measurement but on shuffled careers, and blue lines are predictions from shuffled careers generated by our model. \mathbf{g} - \mathbf{i} , The distribution of the length of streaks P(L) for real careers and $P(L_s)$ for shuffled careers. The hot-streak model (red lines) and its shuffled version (blue lines) closely reproduce P(L) observed in real (red dots) and shuffled careers (blue dots).

and $\log(C_{10})$ for scientists) is drawn from a normal distribution $\mathcal{N}(\Gamma_i, \sigma_i^2)$ that is fixed for an individual. The average Γ_i characterizes the typical impact of works produced by the individual, and σ_i captures the variance. This null model can reproduce the fact that each hit occurs randomly within a career 10,18 . However, it fails to capture any of the temporal correlations observed in Fig. 1. The main reason for this failure is illustrated in Fig. 2a–c, where we selected for illustration purposes one individual from each of the three data sets and measured the dynamics of Γ_i during his or her career. We find that Γ_i is not constant throughout a career. Rather, it deviates from a baseline performance (Γ_0) at a certain point in a career (t_\uparrow) , elevating to a higher value $\Gamma_{\rm H}$ ($\Gamma_{\rm H} > \Gamma_0$), which is then sustained for some time before falling back to level similar to Γ_0 (Fig. 2a–c):

$$\Gamma(t) = \begin{cases} \Gamma_{\mathrm{H}} & t_{\uparrow} \leq t \leq t_{\downarrow} \\ \Gamma_{0} & \text{otherwise} \end{cases}$$
 (1)

This observation, combined with the shortcomings of the null model, raises an intriguing question: could a simple model based on equation (1) explain the temporal anomalies documented in Fig. 1?

To test this hypothesis, we applied equation (1) to real productivity patterns, allowing us to generatively simulate the impacts of the works produced by an individual (Supplementary Information S3.3). During the period in which Γ_H operates, the individual seemingly performs at a higher level than his or her typical performance (Γ_0), prompting us to call this model the hot-streak model (where the Γ_H period corresponds to the hot streak). We introduced to each career one hot streak that occured at random with a fixed duration and magnitude, and repeated our measurements in Fig. 1 on careers generated by the model. We find that, whereas equation (1) introduces only a simple temporal variation,

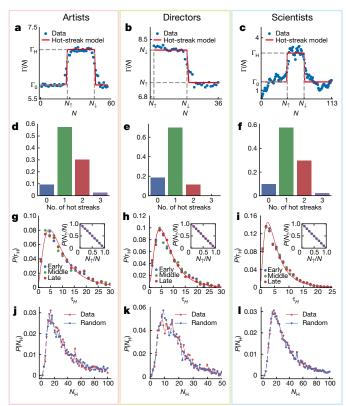


Fig. 2 | **The hot-streak model. a–c**, Γ(N) for one artist (**a**), film director (**b**) and scientist (**c**), selected for illustration purposes. See Extended Data Fig. 9 for randomly selected careers. **d–f**, Histogram of the number of hot streaks in a career. We also measured several performance metrics for individuals who had one or two hot streaks, and found no detectable difference (Extended Data Fig. 10). **g–i**, The distributions of durations of hot streaks $P(\tau_H)$. Red lines are log-normal fits as a visual guide. The insets show cumulative distributions $P\left(\geq \frac{N_{||}}{N} \right)$, indicating that the start of a hot streak $N_{||}$ is distributed randomly among N works in a career. **j–l**, The distributions of the number of works produced during hot streaks $P(N_H)$, compared with a null distribution in which we randomly pick one work as the start of the hot streak. **j**, Artists (n = 3,166). **k**, Directors (n = 5,098). **l**, Scientists (n = 18,121). Two-sided Kolmogorov–Smirnov tests indicate that we cannot reject the hypothesis that the two distributions are drawn from the same distribution (P = 0.12 for artists, P = 0.12 for directors, and P = 0.17 for scientists).

the hot-streak model is sufficient to reproduce all empirical patterns observed in Fig. 1 (Fig. 1d–i and Extended Data Fig. 1s–u). Given the myriad factors that can affect career impacts $^{9-12,18,22,27-30}$, and the obvious diversity of careers we studied, the level of universality and accuracy demonstrated by the simple hot-streak model was unexpected.

The real value of the model arises, however, when we fit the model to real careers to obtain the individual specific Γ_0 , Γ_H , t_\uparrow and t_\downarrow parameters (Supplementary Information 3.4), helping us to reveal several fundamental patterns that govern individual careers.

1. Hot streaks are ubiquitous across careers, yet at the same time usually unique within a career. The vast majority of artists (91%, Fig. 2d), film directors (82%, Fig. 2e) and scientists (90%, Fig. 2f) have at least one hot streak during their careers, documenting the practical relevance of the uncovered hot streak phenomenon. However, despite its ubiquity, the hot streak is likely to be unique within a career. Indeed, when we relaxed our fitting algorithm to allow for multiple hot streaks (up to three) with different values of $\Gamma_{\rm H}$, we found that, among those who had a hot streak, 64% of artists, 80% of directors, and 68% of scientists were best captured by one hot streak only (Fig. 2d–f), documenting the precious nature of hot streaks. Second acts may occur but are less likely, particularly for film directors. Occurrences of more than two hot streaks are rare across all careers.

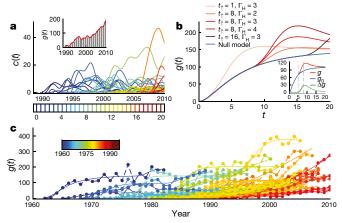


Fig. 3 | The hot streak governs the collective impact of a scientific career. a, Citation patterns of papers published by a randomly selected scientist in our data set. The publication dates are rearranged such that the individual produces a constant number of papers each year (coloured vertical lines, see Methods). Inset, collective impact of the individual, representing the sum of citation dynamics of all of his or her papers. b, Varying hot-streak parameters allows us to reproduce a wide variety of career dynamics (red lines) that cannot be captured by the null model (blue line). For hot-streak parameters in equation (4) (Methods), here we use $\mu = 7.0$, $\sigma = 1.0$, $\Gamma_0 = 1.0$ and $\tau_{\rm H} = 3$ years, but vary $t_{\rm T}$ and $\Gamma_{\rm H}$. Inset, g(t) can be decomposed into $g_0(t)$ and $\Delta g(t)$. c, g(t) of 50 randomly selected scientists. Colour corresponds to a career's starting year, dots denote real data, and solid lines capture the predictions from the hot-streak model.

2. The hot streak occurs randomly within a career. We estimate the beginning of hot streaks, by measuring N_{\uparrow} , the position of work produced when a hot streak starts (t_{\uparrow}) . We find that, across artistic, cultural, and scientific careers, N_{\uparrow} is randomly distributed in the sequence of N works within a career (Fig. 2g–i, insets). This finding reconciles two seemingly divergent schools of thought^{9,10,18}, providing a further explanation for the random impact rule: if the hot streak occurs randomly within a career, and the highest impact works are statistically more likely to appear within a hot streak, then the timing of the highest impact works is also random.

3. Across different domains, hot streaks are considerably shorter than the typical career length recorded in our database. We measure the duration distribution of hot streaks $(\tau_H = t_{\uparrow} - t_{\downarrow})$, finding $P(\tau_H)$ peaks around 5.7 years for artists, 5.2 years for directors, and 3.7 years for scientists, which is largely independent of when it occurs within a career (early, mid or late career; Fig. 2g–i).

4. Unexpectedly, individuals are not more productive during hot streaks. We measured the distribution of the total number of works produced during hot streaks $P(N_{\rm H})$. We then constructed a null distribution, by randomly picking one work in a career and designating its production year to be the start of the hot streak. We found that the two distributions aligned well with each other (Fig. 2j–l). Therefore, individuals show no detectable change in productivity during hot streaks, despite the fact that their outputs in this period are significantly better than the median, suggesting that there is an endogenous shift in individual creativity when the hot streak occurs. For additional properties of hot streaks, see Methods and Extended Data Fig. 5.

To investigate the impact of hot streaks on individual careers, we focused on scientific careers and measured the collective impact of a scientist, g(t), defined as the total number of citations over time collected by all papers published by an individual (Fig. 3a). g(t) can be derived analytically by combining the hot-streak model (equation (1)) and an existing model for the citation patterns of papers (see Methods and Supplementary Information S5), consisting of two terms:

$$g(t) = g_0(t) + \Delta g(t) \tag{2}$$

 $g_0(t)$ captures a career's collective impact in the absence of a hot streak (that is, $\Gamma(t) = \Gamma_0$). Contributions from the hot streak are encoded in $\Delta g(t)$, driven by both the timing and magnitude of hot streaks (see Methods). Varying hot-streak parameters leads to substantial changes in the collective impact of a career (Fig. 3b). Hence the hot-streak model captures a wide range of impact trajectories that are followed by real careers (Fig. 3c), and the accuracy of the model is documented by several metrics (see Methods). Given that individuals improve substantially during hot streaks, the uncovered phenomena can be particularly crucial for understanding the long-term impact of a career (Extended Data Fig. 6).

We further tested several alternative hypotheses, each associated with possible origins of the uncovered hot streaks (see Methods and Supplementary Information S6). Of all hypotheses considered, the hotstreak model is the simplest and least flexible. However, it is the only model whose predictions are consistent with real careers (Extended Data Figs 7, 8). The identification of the true origins of hot streaks is beyond the scope of this work. As such, the hot streaks uncovered here should be treated in a metaphorical sense, highlighting an intriguing period of outstanding performance during individual careers without implying any associated drivers. Crucially, though, the findings presented here hold the same, regardless of the underlying drivers.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0315-8.

Received: 20 October 2017; Accepted: 4 June 2018; Published online 11 July 2018.

- Gilovich, T., Vallone, R. & Tversky, A. The hot hand in basketball: on the misperception of random sequences. Cognit. Psychol. 17, 295–314 (1985)
- Miller, J. B. & Sanjurjo, A. Surprised by the gambler's and hot hand fallacies? A truth in the law of small numbers. *IGIER Working Paper No.* 552 (2016).
- Ayton, P. & Fischer, I. The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? Mem. Cognit. 32, 1369–1378 (2004).
- Rabin, M. & Vayanos, D. The gambler's and hot-hand fallacies: theory and applications. Rev. Econ. Stud. 77, 730–778 (2010).
- Xu, J. & Harvey, N. Carry on winning: the gamblers' fallacy creates hot hand effects in online gambling. Cognition 131, 173–180 (2014).
- Hendricks, D., Patel, J. & Zeckhauser, R. Hot hands in mutual funds: short-run persistence of relative performance, 1974–1988. J. Finance 48, 93–130 (1993).
- Kahneman, D. & Riepe, M. W. Aspects of investor psychology. J. Portfol. Manage. 24, 52–65 (1998).
- 8. Lehman, H. C. Age and Achievement (Princeton Univ. Press, Princeton, 1953).
- 9. Merton, R. K. The Matthew effect in science. Science 159, 56–63 (1968).
- Simonton, D. K. Age and outstanding achievement: what do we know after a century of research? Psychol. Bull. 104, 251–267 (1988).
- 11. Jones, B. F. The burden of knowledge and the death of the renaissance man: is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).
- Petersen, A. M. et al. Reputation and impact in academic careers. Proc. Natl Acad. Sci. USA 111, 15316–15321 (2014).
- Stringer, M. J., Sales-Pardo, M. & Nunes Amaral, L. A. Effectiveness of journal ranking schemes as a tool for locating information. PLoS One 3, e1683 (2008).
- Radicchi, F., Fortunato, S. & Castellano, C. Universality of citation distributions: toward an objective measure of scientific impact. *Proc. Natl Acad. Sci. USA* 105, 17268–17272 (2008).
- Galenson, D. W. Old Masters and Young Geniuses: The Two Life Cycles of Artistic Creativity (Princeton Univ. Press, Princeton, 2011).
- Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. Science 342, 127–132 (2013).
- Way, S. F., Morgan, A. C., Clauset, A. & Larremore, D. B. The misleading narrative of the canonical faculty productivity trajectory. *Proc. Natl Acad. Sci. USA* 114, E9216–E9223 (2017).
- Sinatra, R., Wang, D., Deville, P., Song, C. & Barabási, A.-L. Quantifying the evolution of individual scientific impact. Science 354, aaf5239 (2016).
- 19. Fortunato, S. et al. Science of science. *Science* **359**, eaao0185 (2018).
- Zeng, A. et al. The science of science: From the perspective of complex systems. Phys. Rep. 714–715, 1–73 (2017).
- Simonton, D. K. Scientific Genius: A Psychology of Science (Cambridge Univ. Press, Cambridge, 1988).
- Duch, J. et al. The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. PLoS One 7, e51332 (2012).
- Price, D. S. A general theory of bibliometric and other cumulative advantage processes. J. Assoc. Inf. Sci. Technol. 27, 292–306 (1976).



- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. Science 286, 509–512 (1999).
- Wasserman, M., Zeng, X. H. T. & Amaral, L. A. N. Cross-evaluation of metrics to estimate the significance of creative works. *Proc. Natl Acad. Sci. USA* 112, 1281–1286 (2015).
- 26. Uzzi, B., Mukherjeé, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
- 27. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
- 28. Shockley, W. On the statistics of individual variations of productivity in research laboratories. *Proc. IRE* **45**, 279–290 (1957).
- Redner, S. Citation statistics from 110 years of physical review. Phys. Today 58, 49 (2005).
- Moreira, J. A., Zeng, X. H. T. & Amaral, L. A. N. The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *PLoS One* 10, e0143108 (2015).

Acknowledgements We thank A.-L. Barabasi, R. Burt, J. Chown, J. Evans, C. Jin, L. Nordgren, W. Ocasio, B. Uzzi, Y. Yin, the Kellogg Insights and all members of NICO for invaluable comments. This work is supported by the Air Force Office of Scientific Research (AFOSR) under award number FA9550-15-1-0162 and FA9550-17-1-0089, and Northwestern University's Data Science Initiative. R.S. acknowledges support from AFOSR grant

FA9550-15-1-0364 and from the Central European University Intellectual Themes Initiative 'Just Data'.

Reviewer information *Nature* thanks J. Walsh, J. West and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.W. conceived the project and designed the experiments; L.L. and Y.W. collected data and performed empirical analyses with help from R.S., C.L.G., C.S. and D.W.; L.L., Y.W., C.S. and D.W. did theoretical calculations; all authors discussed and interpreted results; D.W. and L.L. wrote the manuscript; all authors edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0315-8.

 $\label{lem:condition} \textbf{Supplementary information} \ is \ available \ for \ this \ paper \ at \ https://doi.org/10.1038/s41586-018-0315-8.$

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.W. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data description. We compiled three large-scale data sets of individual careers across three major domains involving human creativity. The first data set (D_1) consists of auction records curated from online auction databases, allowing us to reconstruct the career histories of 3,480 artists through the sequence of works they each produced, together with the impacts of the artworks, approximated by hammer prices in auctions¹⁵. D_2 contains profiles of 6,233 film directors recorded in the IMDB database, each career being represented by the sequence of films directed by the individual. As metrics that quantify the impact of a film correlate closely with each other²⁵, here we use the IMDB rating to measure the goodness of a film. Finally, our third data set (D_3) includes publication records of 20,040 individual scientists through a large-scale name disambiguation effort that combined the Web of Science (https://clarivate.com/products/web-of-science/) and Google Scholar (https://scholar.google.co.uk/) data sets. The impact of each paper is measured by citations garnered after 10 years of its publication 13,16,18,26 (C_{10}) . Further details on data collection and curation are provided in Supplementary Information S1.

To study the impact of works across the three domains, we measured the distributions of hammer prices, IMDB ratings and paper citation counts in our data sets. Both hammer price (D_1) and C_{10} (D_3) follow fat-tailed distributions, well approximated by a log-normal function (Extended Data Fig. 2a, c), and the IMDB rating follows a normal distribution ranging between 1 and 10 (Extended Data Fig. 2b). To make sure C_{10} is not affected by citation inflation 14,18,31 , we also measured a rescaled C_{10} (see Supplementary Information S1.3) and found that it also followed a fat-tailed distribution (Extended Data Fig. 2c, inset). Therefore, we take the logarithmic of hammer price and C_{10} (log(price) and log(C_{10})) to approximate the goodness of an artwork and scientific publication. Note that the choice of logarithmic for hammer prices and C_{10} is meant to be consistent with prior studies ^{14,18}, and does not affect any of the conclusions of the paper. Indeed, the logarithmic function is a monotonically increasing function, hence it does not change the rank ordering of top hits in a career. Note that while the data sets we used in this paper cover a large collection of career histories across a wide range of domains, the data-driven nature of our study indicates that the scope of our data is limited to individuals who have had sufficiently long careers to provide enough data points for statistical analyses (Supplementary Information S1).

Additional properties of hot streaks. How much does an individual deviate from his or her typical performance during a hot streak? Do people with higher Γ_0 also experience more performance gain from hot streaks? We explored correlations between Γ_0 and Γ_H , finding them to be well approximated by a linear relationship (Extended Data Fig. 5a–c). Hence, individuals with better typical performance also perform better during their hot streaks. It is interesting to note that the coefficients are slightly less than 1 (Extended Data Fig. 5a–c). Hence $\Delta\Gamma \equiv \Gamma_H - \Gamma_0$ decreases with Γ_0 (Extended Data Fig. 5a–c, insets), suggesting that individuals with smaller Γ_0 benefit more from hot streaks. These results are again independent of when the hot streak occurs along a career (Extended Data Fig. 5a–c).

The temporally localized nature of the hot streak is also captured by its proportion over career length $\tau_{\rm H}/T$ (Extended Data Fig. 5d–f). We compared the duration of hot streaks with typical career length, finding that the median hovers around 20% (0.17 for artists, 0.23 for directors, and 0.20 for scientists).

Analytical solutions for the collective impact of a scientific career, g(t). Brought into the spotlight by popular websites such as Google Scholar, g(t) is playing an increasingly important role in driving many critical decisions, from hiring, promotion and tenure to awarding of grants and rewards. Many factors are known to influence it, ranging from productivity^{17,28} to citation disparity and dynamics^{13,14,16,22,23,29} and temporal inhomogeneities along a career^{11,17,18,21,22,30}. As our goal is to understand impact, here we bypass the need to evaluate the inhomogeneous nature of productivity^{17,18} by rearranging the publication time of each paper, such that an individual produces a constant number of papers each year, denoted by n (Fig. 3a). To calculate g(t), we need to incorporate the citations patterns of papers into our hot-streak model (equation (1)). A recent study¹⁶ suggested that the citation dynamics of a paper published at time t_0 can be approximated by

$$C(t,t_0) = m \left[e^{\lambda \Phi \left(\frac{\ln(t - t_0) - \mu}{\sigma} \right)} - 1 \right] \equiv m \left[e^{\Gamma(t_0) \Phi \left(\frac{\ln(t - t_0) - \mu}{\sigma} \right)} - 1 \right], \tag{3}$$

where m is a global parameter describing the typical number of references a paper contains, and $\Phi(\cdot)$ is the cumulative normal function, characterized by μ and σ , which capture the typical citation life cycle of a paper. The paper's ultimate impact is determined by its fitness¹⁶, λ . To adapt equation (3) into our framework, we replace λ with $\Gamma(t_0)$, and for simplicity assume that μ and σ are fixed for different papers published by an individual. The resulting model, combining equations (1) and (3), can be solved analytically (Supplementary Information S5), allowing us to express g(t) in terms of hot-streak parameters:

$$g(t) = \underline{nm} \left[e^{\Gamma_0 \Phi \left(\frac{\ln(t) - \mu}{\sigma} \right)} - 1 \right]$$

$$+ \left[\frac{0}{nm(\Gamma_{H} - \Gamma_0) \Phi \left(\frac{\ln(t - t_{\uparrow}) - \mu}{\sigma} \right) C(t, t_{\uparrow})} t_{\uparrow} \leq t < t_{\downarrow}}{m(\Gamma_{H} - \Gamma_0) \left[\Phi \left(\frac{\ln(t - t_{\uparrow}) - \mu}{\sigma} \right) C(t, t_{\uparrow}) - \Phi \left(\frac{\ln(t - t_{\downarrow}) - \mu}{\sigma} \right) C(t, t_{\downarrow}) \right]} t \geq t_{\downarrow}$$

$$+ \underbrace{\left[\frac{nm(\Gamma_{H} - \Gamma_0) \Phi \left(\frac{\ln(t - t_{\downarrow}) - \mu}{\sigma} \right) C(t, t_{\downarrow})}{\sigma} \right] t \geq t_{\downarrow}}_{\Delta g(t)}$$

$$(4)$$

Equation (4) consists of two terms. $g_0(t)$ captures a career's collective impact in the absence of hot streaks (that is, $\Gamma(t) = \Gamma_0$). Contributions from the hot streak are encoded in $\Delta g(t)$, driven by both the timing and magnitude of hot streaks $(t_{\hat{\Gamma}}, t_{\hat{\Gamma}}, \Gamma_{\hat{\Pi}}, \text{and } \Gamma_{\hat{\Pi}} - \Gamma_0)$.

Evaluating the accuracy of the hot-streak model. We quantify the accuracy of our model in equation (4) using three metrics.

To account for the inherently noisy career trajectories, we first assign an impact envelope to each career, explicitly quantifying the uncertainty of model predictions (Extended Data Fig. 6g). We simulated g(t) for each individual by assigning a Gaussian noise $\mathcal{N}(0, \sigma_s^2)$ to the fitted Γ_0 and Γ_H . For each paper i we randomly draw its Γ_i from a normal distribution, depending on whether the paper was published within the hot streak (Γ_H during hot streak, Γ_0 otherwise). The standard deviation σ_s represents the inherent noise of the goodness parameter defined in Supplementary Information S3.5. For each individual, we simulated g(t) for 1,000 realizations, allowing us to obtain a distribution of g(t), with one standard deviation offering an uncertainty envelope. We repeated the same procedures for the null model. We measure the fraction of g(t) that falls within the envelope, finding that the distribution peaks close to 1 (Extended Data Fig. 6h), which indicates that most career trajectories are well encapsulated within the predicted envelopes.

The superior accuracy of our model is also captured by the mean absolute percentage error (MAPE). We compared the distribution of MAPE between the data and the predictions of the model (Extended Data Fig. 6f), finding again that the hot-streak model outperformed the null model. The improvement was most pronounced for an early onset of hot streaks (Extended Data Fig. 6i), which is also consistent with our model's predictions.

To account for model complexity, we also calculated the Bayesian information criterion (BIC) measure, which penalizes the number of parameters in the model. Compared with the null model, the hot-streak model has systematically smaller BIC (Extended Data Figs 6e), documenting that the hot-streak model better captures the collective impact of a career than the null model.

Implications of hot streaks for long-term career impact. The analytical framework presented here not only offers a new theoretical basis for our quantitative understanding of dynamical patterns governing individual career impact, but also may have implications for comparing and evaluating scientists (Extended Data Fig. 6). Indeed, for individuals whose hot streaks are yet to come, ignoring the hot streak may lead to underestimation of their impacts (Extended Data Fig. 6a, b), especially given the ubiquitous nature of hot streaks (Fig. 2f). On the other hand, an early onset of a hot streak leads to a high impact that peaks early but may not be sustained unless a second hot streak occurs (Extended Data Fig. 6c). Testing alternative hypotheses. To explore the possibility that alternative hypotheses might explain the observed patterns, we tested several models that capture different dynamics of hot streaks (Supplementary Information S6.3), each associated with possible origins of the uncovered hot streaks. (A) A right trapezoid (Extended Data Fig. 7b) captures a sudden onset of a hot streak with a more gradual decline, as innovators may stumble upon a groundbreaking idea, which manifests itself in the forms of multiple artworks, films, or publications. Hence from an evolutionary perspective, the duration of a hot streak may characterize the time it takes for the temporary competitive advantage to dissipate. (B) An isosceles trapezoid model (Extended Data Fig. 7c) captures hot streaks that evolve and dissolve gradually over time, which may approximate social tie dynamics, as one individual's hot streak could be the result of a fruitful, repeated collaboration $^{27,32}. \ (C)$ Furthermore, individual performance may peak at a certain point in a career, prompting us to test inverted-U shape (Extended Data Fig. 7d) and tent functions (Extended Data Fig. 7e). (D) Last, a left trapezoid function (Extended Data Fig. 7f) captures a



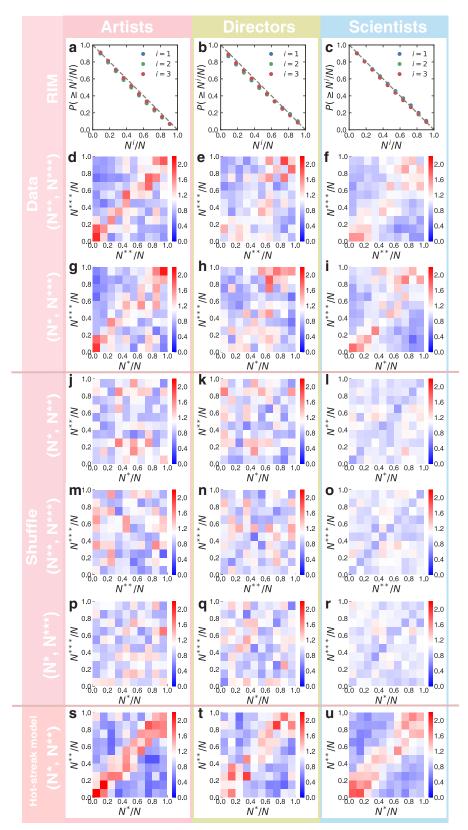
gradual startup period with a sharp cutoff, corresponding to career opportunities that can augment impact but last for a fixed duration.

We tested the validity of the four alternative hypotheses (A–D) by comparing each model's prediction with empirical observations on the relative order of the top six hits within a career. The symmetric patterns of ϕ and $R\left(\frac{\Delta N}{N}\right)$ observed in real careers suggest that the biggest hit is equally likely to appear before or after the second-biggest hit. The randomness of the relative ordering among hits is not limited to the two biggest hits. Indeed, we measured the position of the top three hits $(\tilde{N}:\text{mover})$ relative to the top six hits of the career, and compute $P(\tilde{N})$ for each of the three hits for artists, directors and scientists. We found a lack of predictive patterns for $P(\tilde{N})$ across the three domains, suggesting that the relative orders among the top six hits in real careers are random (Extended Data Fig. 7g, o, w). We tested hypotheses A–D systematically to describe real careers (Supplementary Information S6.3), and found that the hot-streak model was the only model whose predictions were consistent with real careers (Extended Data Fig. 7h–m, p–u, x–ac). As such, the hot-streak model also offers a superior fit to the data than the other models (Extended Data Fig. 7n, v, ad).

We also tested whether Markov models could account for our observations (Supplementary Information S6.2). We explored multiple variants of Markov models by introducing short-range correlations between the impacts of adjacent works, correlations between the volatility of their impacts, and hidden Markov model with two states, finding again that the hot-streak model stood out in its ability to describe the observed patterns (Extended Data Fig. 8 and Supplementary Information S6.2). Together, these results demonstrate that none of these alternative hypotheses alone can account for the empirical observations in real careers.

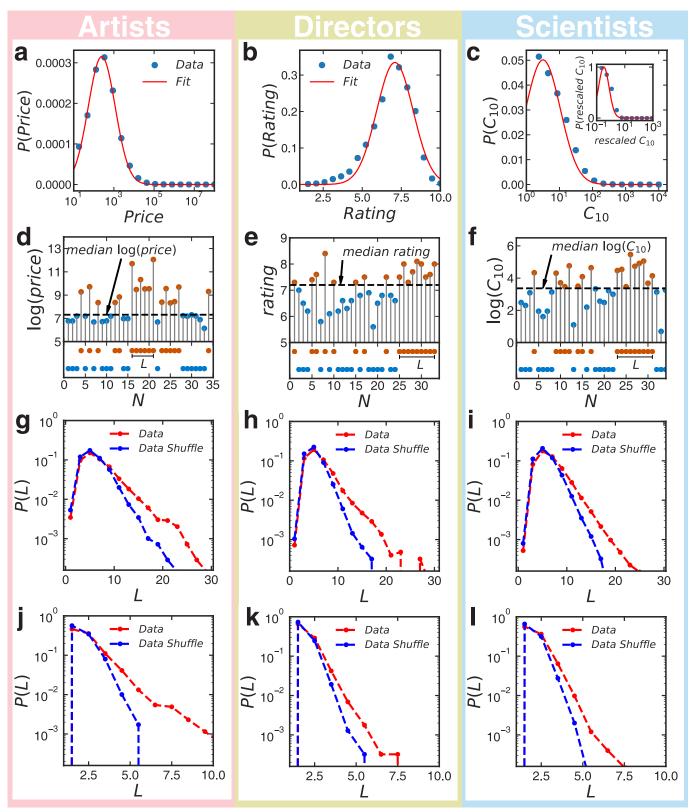
Code availability. Code is available at https://lu-liu.github.io/hotstreaks/. **Data availability.** The data are available at https://lu-liu.github.io/hotstreaks/.

- Pan, R. K., Petersen, A. M., Pammolli, F. & Fortunato, S. The memory of science: inflation, myopia, and the knowledge network. Preprint at https://arxiv.org/abs/ 1607.05606 (2016).
- Palla, G., Barabási, Á.-L. & Vicsek, T. Quantifying social group evolution. Nature 446, 664–667 (2007).



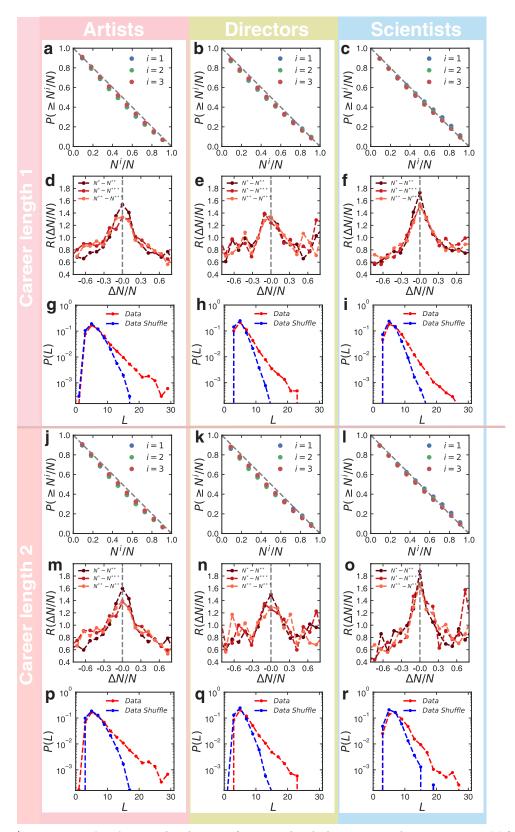
Extended Data Fig. 1 | Additional results on hot streaks in artistic, cultural, and scientific careers. \mathbf{a} – \mathbf{c} , The cumulative distribution $P(\geq N^i/N)$ for the order of the top three highest impact works within a career for artists (\mathbf{a}), directors (\mathbf{b}) and scientists (\mathbf{c}). N^i denotes the order of the i^{th} highest-impact work within a career. The colours denote different hit works, and the dashed grey line denotes $P(\geq N^i/N)$ for a uniform distribution. \mathbf{d} – \mathbf{f} , $\phi(N^{***}, N^{****})$ for the second- and third-highest-impact works within a career. $\phi(N^{**}, N^{****})$ is also overrepresented along the

diagonal. \mathbf{g} — \mathbf{i} , $\phi(N^*, N^{***})$ for the first- and third-highest-impact works within a career. \mathbf{j} — \mathbf{r} , We shuffled the order of each work in a career while keeping their impact intact. The diagonal patterns in \mathbf{d} — \mathbf{i} and Fig. 1a—c disappeared for shuffled careers. \mathbf{s} — \mathbf{u} , $\phi(N^*, N^{**})$ predicted by the hotstreak model successfully recovered the diagonal patterns observed in \mathbf{a} — \mathbf{c} . For \mathbf{d} — \mathbf{u} and Fig. 1a—c, we applied the same binning procedure to data, using bins that ranged from 0 to 1 with increments of 0.1.



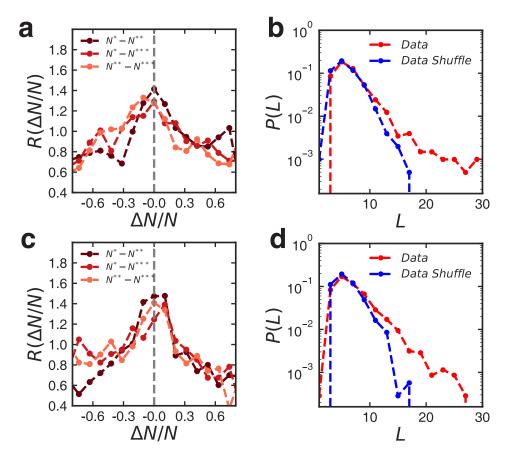
Extended Data Fig. 2 | Measuring the length of streaks using different thresholds. a, The distribution of auction price P(Price) for artists. Blue dots denote data, and the red line is a log-normal distribution with average $\mu=7.9$ and standard deviation $\sigma=1.5$. b, The distribution of film rating P(Rating) for directors. The red line is a normal distribution with average $\mu=7.1$ and standard deviation $\sigma=1.2$. c, The distribution of raw and rescaled C_{10} (inset) for scientists. The red line is a log-normal distribution, with $\mu=2.3$ and $\sigma=1.3$ for c and $\mu=-0.4$ and $\sigma=0.8$ for the inset.

d–**f**, Definitions of the longest streak L for artists (**d**), directors (**e**) and scientists (**f**). Dots are coloured orange above the threshold, blue otherwise. The lower panel highlights the longest streak in a career. \mathbf{g} –**i**, P(L) for real careers and $P(L_s)$ for shuffled careers using the mean impact within a career as the threshold. \mathbf{j} –**l**, As in \mathbf{g} –**i**, but using the top 10% impact as the threshold to calculate L and L_s . In all cases, P(L) has a wider tail than $P(L_s)$, indicating that high-impact works in real careers tend to cluster together.



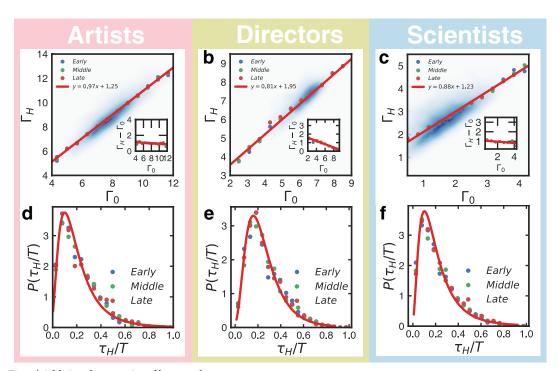
Extended Data Fig. 3 | **Varying career length.** To test the robustness of our results, we repeated our measurements by controlling for the career length of individuals. \mathbf{a} - \mathbf{i} , Artists and directors with careers of at least 20 years and scientists with careers of at least 30 years. \mathbf{a} - \mathbf{c} , $P(\geq N^i/N)$ of the top three highest-impact works within a career. \mathbf{d} - \mathbf{f} , $R(\frac{\Delta N}{N})$ among the

top three highest-impact works in a career. \mathbf{g} – \mathbf{i} , P(L) for real careers and $P(L_{\rm s})$ for shuffled careers. \mathbf{j} – \mathbf{r} , As in \mathbf{a} – \mathbf{i} but for artists and directors with careers of at least 30 years and scientists with careers of at least 40 years. These results demonstrate that the patterns observed in Fig. 1 hold for individuals with different career lengths.



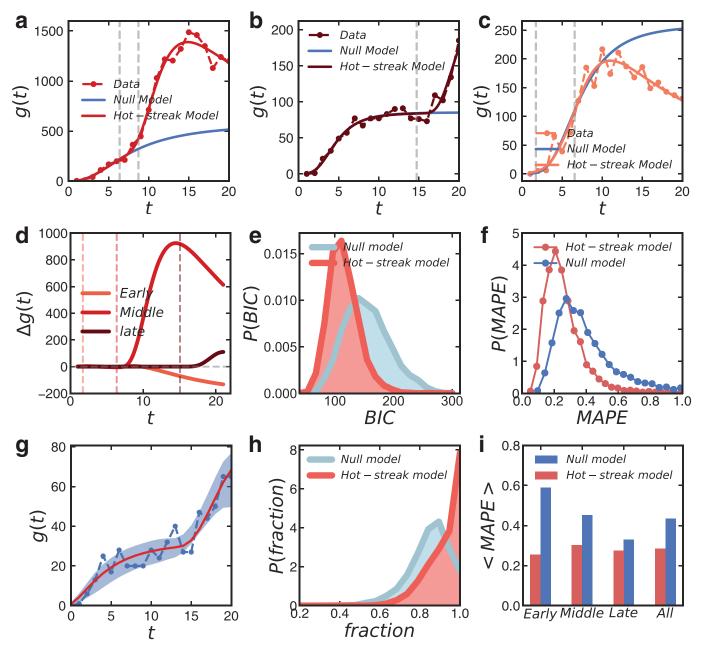
Extended Data Fig. 4 | **Artistic careers from different eras. a**, $R\left(\frac{\Delta N}{N}\right)$ for artists who started their careers before 1850. **b**, P(L) for real careers and $P(L_s)$ for shuffled careers for artists who started their careers before 1850. **c**, $R\left(\frac{\Delta N}{N}\right)$ for artists who started careers between 1850 and 1900. **d**, P(L) for

real careers and $P(L_s)$ for shuffled careers for artists who started their careers between 1850 and 1900. These results demonstrate that the patterns observed in Fig. 1 hold for artists from different eras.



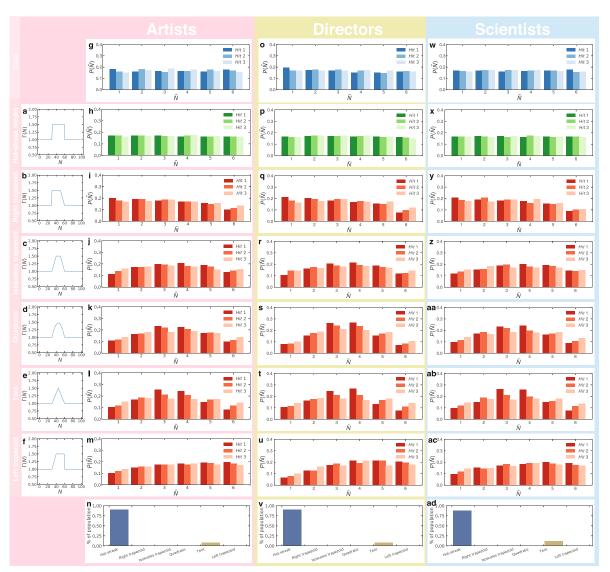
Extended Data Fig. 5 | Additional properties of hot streaks. a-c, Correlations between $\Gamma_{\rm H}$ and Γ_0 for artists (a; n=3,166), directors (b; n=5,098) and scientists (c; n=18,121). The blue background denotes the kernel density of data, dots represent binning results of data, and the red lines depict a linear fit. Inset, the relationship between

 $\Delta\Gamma$ ($=\Gamma_{\rm H}-\Gamma_0)$ and Γ_0 . **d-f**, The distribution of $\tau_{\rm H}/T$, representing the duration of hot streaks over total career lengths. The temporally localized nature of a hot streak is also captured by its proportion over career length $\tau_{\rm H}/T$.



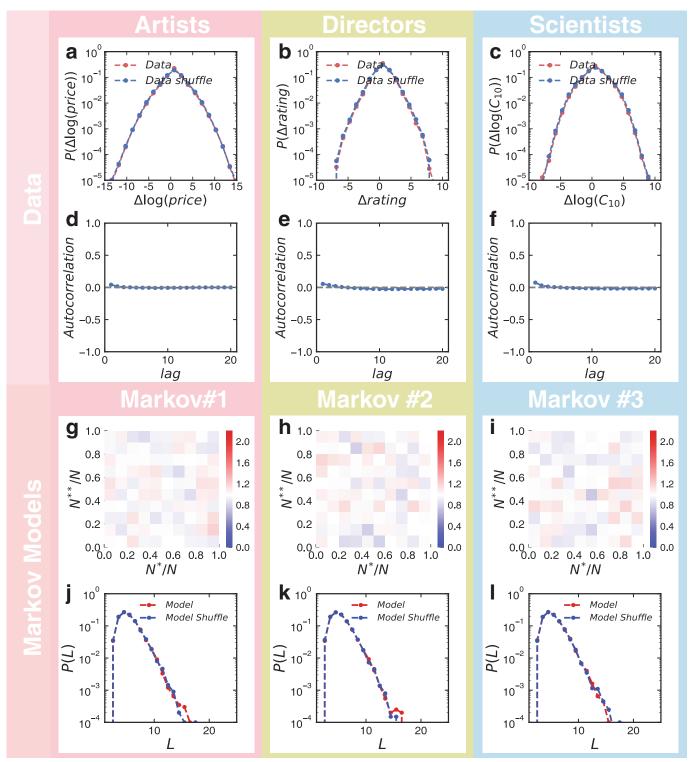
Extended Data Fig. 6 | Comparison of g(t) between the null model and the hot-streak model. a-c, g(t) of three scientists in our data set with midcareer (a), late-career (b) and early-career (c) onset of hot streaks. Red dots denote data, the blue line is the null model's prediction based on early performance, and the red line captures the predictions from the hot-streak model, with dashed grey lines denoting the start and end of hot streaks. d, The difference between our hot-streak model and the null model for each individual, $\Delta g(t)$. Dashed lines with corresponding colours denote the start of the hot streak. d illustrates the discrepancies in estimating an individual's future impact if we ignore the uncovered hot streaks. e, The distribution of the BIC measure, P(BIC), showing that the hot-streak model outperforms the null model in describing g(t) after accounting for

model complexity. **f**, The distribution of the MAPE measure, P(MAPE), showing that the hot-streak model outperforms the null model in describing g(t). **g**, The uncertainty envelope of g(t) for an individual in our data set. Blue dots denote data, and the red line is the fitting result of equation (4). Shaded area illustrates predicted uncertainty (one standard deviation). **h**, The fraction of g(t) falling within the envelope for the null model (blue) and our hot-streak model (red). Fraction = 1.0 indicates that the entire g(t) trajectory falls within the envelope. **i**, Average MAPE of our hot-streak model and the null model for individuals with early-career, mid-career and late-career onset of hot streaks. The difference is largest for individuals with early-onset hot streaks and smallest for those with late-onset ones.



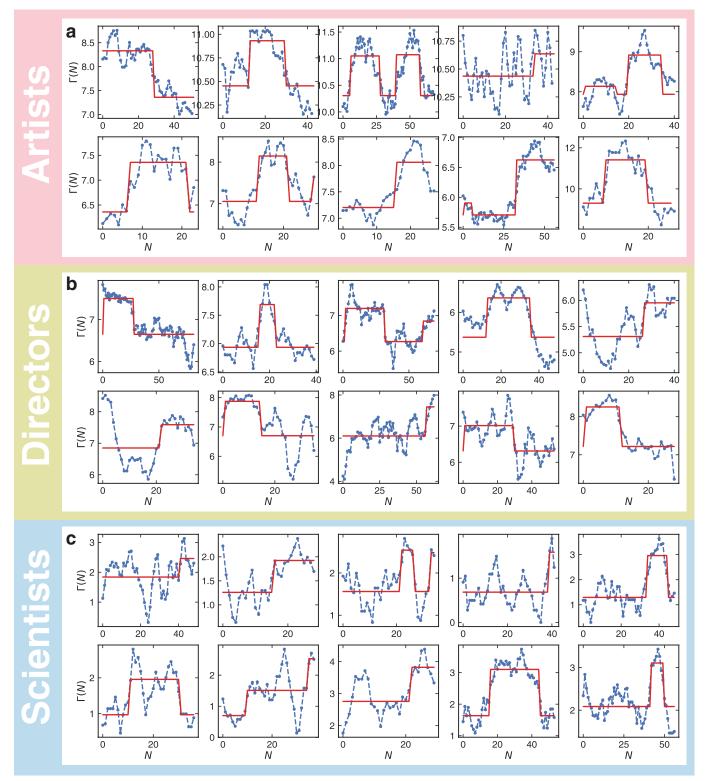
Extended Data Fig. 7 | Testing alternative hot-streak dynamics. a-f, Illustrative examples of $\Gamma(N)$ for the hot-streak model (a), right trapezoid function (b), isosceles trapezoid function (c), quadratic function (d), tent function (e) and left trapezoid function (f). g, The distribution of the relative position $P(\tilde{N})$ of the three highest-impact works among the six highest-impact works within a career for artists, where \tilde{N} denotes the relative order among the top six hits. h-m, $P(\tilde{N})$ predicted by corresponding models shown in a-f, respectively, according to artists' real productivity profiles. To test whether data agree with model predictions, we measured their statistical difference using the P value of the Kolmogorov–Smirnov test for discrete distributions. We colour the

distributions green if we cannot reject the hypothesis that the data and the model predictions come from the same distributions, and red otherwise. Among the six models, the hot-streak model is the only model whose predictions are consistent with the data in terms of the relative ordering among the six highest-impact works observed in real careers. **n**, The proportion of real careers that are captured by the model with the smallest BIC among different hypotheses. The hot-streak model again stands out as the best model to describe real careers. We repeated the analyses for directors (**p**-**v**) and scientists (**x**-**ad**), the conclusions remained the same across all three domains.



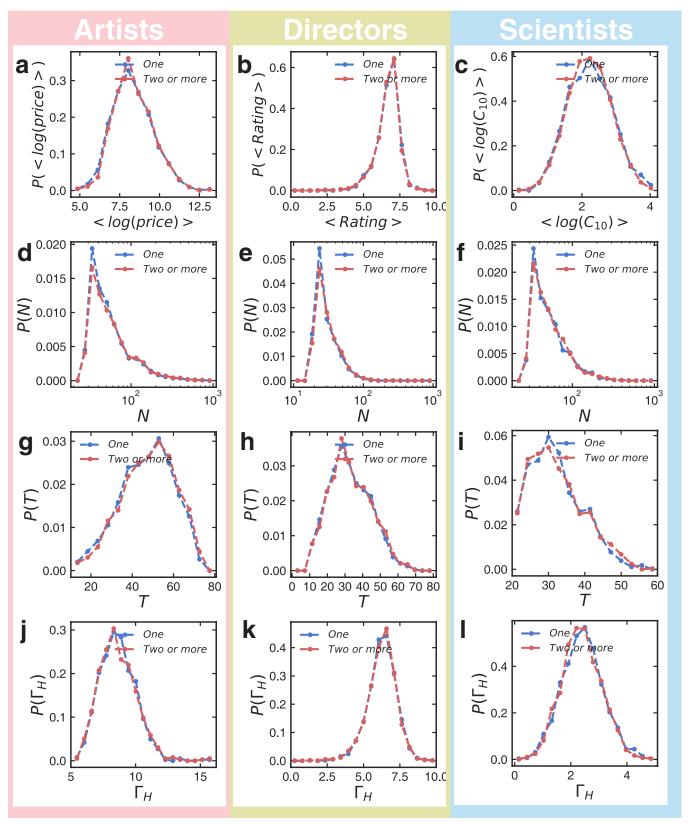
Extended Data Fig. 8 | **Testing Markovian hypotheses.** Here we test whether the observed patterns can be explained by Markovian dynamics that introduce correlations between neighbouring data points. We first test the assumptions of the Markovian hypothesis from the data (\mathbf{a} – \mathbf{f}). \mathbf{a} – \mathbf{c} , The distribution of N, N+1 differences between adjacent data points observed in real careers for artists (\mathbf{a} , n=3,480), directors (\mathbf{b} , n=6,233) and scientists (\mathbf{c} , n=20,040). \mathbf{d} – \mathbf{f} , The autocorrelation measured in real careers for artists (\mathbf{d} , n=3,480), directors (\mathbf{e} , n=6,233), and scientists (\mathbf{f} , n=20,040). \mathbf{a} – \mathbf{f} suggest that there is little short-range correlation

in data across the three domains. We test three variants of Markovian models (g–l). The details of these models are outlined in Supplementary Information S6.2. g–i, $\phi(N^*, N^{**})$ of the top two highest-impact works within a career for three Markovian models using scientists' profiles as input. j–l, The distribution of the longest streak length P(L) and $P(L_s)$ using median impact within a career as threshold for the three Markovian models. g–l demonstrate that the three Markovian models failed to capture the observed colocations among hits.



Extended Data Fig. 9 | **Additional examples of** Γ **. a–c**, Each subplot denotes the fitting result on Γ sequence for a randomly selected career for artists (**a**), directors (**b**) and scientists (**c**). Blue dots denote the moving

average $\Gamma(N)$ from data and red lines denote the best fitting result of the hot-streak model for each individual.



Extended Data Fig. 10 | Individuals with one or more hot streaks. $\mathbf{a}-\mathbf{c}$, The distribution of average impacts for individuals with one or more than one hot streaks for artists (\mathbf{a}), directors (\mathbf{b}) and scientists (\mathbf{c}). Blue dots denote individuals with one hot streak, and red dots denote individuals with at least two hot streaks. $\mathbf{d}-\mathbf{f}$, The distribution of the number of works P(N) within a career for individuals with one or more than one hot streak for for artists (\mathbf{d}), directors (\mathbf{e}) and scientists (\mathbf{f}). $\mathbf{g}-\mathbf{i}$, The distribution of career length $P(\tau)$ for individuals with one or more

than one hot streaks for artists (g), directors (h) and scientists (i). j–l, The distribution of $P(\Gamma_{\rm H})$ for individuals with one or more than one hot streaks for artists (j), directors (k) and scientists (l). Between those who have one or two hot streaks, there is no detectable difference in terms of typical performance metrics, including impact, productivity and career length, suggesting that the hot streak captures an orthogonal dimension to current metrics characterizing individual careers.



Prediction of acute myeloid leukaemia risk in healthy individuals

Sagi Abelson^{1,46}, Grace Collord^{2,3,46}, Stanley W. K. Ng⁴, Omer Weissbrod⁵, Netta Mendelson Cohen⁵, Elisabeth Niemeyer⁶, Noam Barda⁷, Philip C. Zuzarte⁸, Lawrence Heisler⁸, Yogi Sundaravadanam⁸, Robert Luben⁹, Shabina Hayat⁹, Ting Ting Wang^{1,10}, Zhen Zhao¹, Iulia Cirlan¹, Trevor J. Pugh^{1,8,10}, David Soave⁸, Karen Ng⁸, Calli Latimer², Claire Hardy², Keiran Raine², David Jones², Diana Hoult¹¹, Abigail Britten¹¹, John D. McPherson⁸, Mattias Johansson¹², Faridah Mbabaali⁸, Jenna Eagles⁸, Jessica K. Miller⁸, Danielle Pasternack⁸, Lee Timms⁸, Paul Krzyzanowski⁸, Philip Awadalla⁸, Rui Costa¹³, Eran Segal⁵, Scott V. Bratman^{1,8,14}, Philip Beer², Sam Behjati^{2,3}, Inigo Martincorena², Jean C. Y. Wang^{1,15,16}, Kristian M. Bowles^{17,18}, J. Ramón Quirós¹⁹, Anna Karakatsani^{20,21}, Carlo La Vecchia^{20,22}, Antonia Trichopoulou²⁰, Elena Salamanca–Fernández^{23,24}, José M. Huerta^{24,25}, Aurelio Barricarte^{24,26,27}, Ruth C. Travis²⁸, Rosario Tumino²⁹, Giovanna Masala³⁰, Heiner Boeing³¹, Salvatore Panico³², Rudolf Kaaks³³, Alwin Krämer³⁴, Sabina Sieri³⁵, Elio Riboli³⁶, Paolo Vineis³⁶, Matthieu Foll¹², James McKay¹², Silvia Polidoro³⁷, Núria Sala³⁸, Kay–Tee Khaw³⁹, Roel Vermeulen⁴⁰, Peter J. Campbell^{2,41}, Elli Papaemmanuil^{2,42}, Mark D. Minden^{1,10,15,16}, Amos Tanay⁵, Ran D. Balicer⁷, Nicholas J. Wareham¹¹, Moritz Gerstung^{2,13,47*}, John E. Dick^{1,43,47*}, Paul Brennan^{12,47*}, George S. Vassiliou^{2,41,44,47*} & Liran I. Shlush^{1,6,45,47*}

The incidence of acute myeloid leukaemia (AML) increases with age and mortality exceeds 90% when diagnosed after age 65. Most cases arise without any detectable early symptoms and patients usually present with the acute complications of bone marrow failure¹. The onset of such de novo AML cases is typically preceded by the accumulation of somatic mutations in preleukaemic haematopoietic stem and progenitor cells (HSPCs) that undergo clonal expansion^{2,3}. However, recurrent AML mutations also accumulate in HSPCs during ageing of healthy individuals who do not develop AML, a phenomenon referred to as age-related clonal haematopoiesis (ARCH)⁴⁻⁸. Here we use deep sequencing to analyse genes that are recurrently mutated in AML to distinguish between individuals who have a high risk of developing AML and those with benign ARCH. We analysed peripheral blood cells from 95 individuals that were obtained on average 6.3 years before AML diagnosis (pre-AML group), together with 414 unselected age- and gendermatched individuals (control group). Pre-AML cases were distinct from controls and had more mutations per sample, higher variant allele frequencies, indicating greater clonal expansion, and showed enrichment of mutations in specific genes. Genetic parameters were used to derive a model that accurately predicted AML-free survival; this model was validated in an independent cohort of 29 pre-AML cases and 262 controls. Because AML is rare, we also developed an AML predictive model using a large electronic health record database that identified individuals at greater risk. Collectively our findings provide proof-of-concept that it is possible to discriminate ARCH from pre-AML many years before malignant transformation. This could in future enable earlier detection and monitoring, and may help to inform intervention.

To examine the occurrence of somatic mutations before the development of AML, we carried out deep error-corrected targeted sequencing of AML-associated genes in a discovery cohort of 95 pre-AML cases and 414 age- and gender-matched controls (Supplementary Table 1). A validation cohort comprising 29 pre-AML cases and 262 controls (Supplementary Table 1) was analysed using deep sequencing with an overlapping gene panel. Taking both cohorts together, ARCH, defined on the basis of putative driver mutations (ARCH-PD), was found in 73.4% of the pre-AML cases at a median of 7.6 years before diagnosis. By contrast, ARCH-PD was observed in 36.7% of controls ($P < 2.2 \times 10^{-16}$, two-sided Fisher's exact test; Fig. 1a), consistent with data from a study of more than 2,000 unselected individuals assayed using a similarly sensitive method^{9,10}. Additionally, 39% of pre-AML cases above the age of 50 had a driver mutation with a variant allele frequency (VAF) of more than 10%, compared to only 4% of controls,

1 Princess Margaret Cancer Centre, University Health Network (UHN), Toronto, Ontario, Canada. 2 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. 3 Department of Paediatrics, University of Cambridge, Cambridge, UK, 4Institute of Biomaterials and Biomedical Engineering, University of Toronto, Ontario, Canada, 5Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. 6Department of Immunology, Weizmann Institute of Science, Rehovot, Israel. 7Clalit Research Institute, Tel Aviv, Israel. 8Ontario Institute for Cancer Research, Toronto, Ontario, Canada. 9Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK. 10 Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. 11 MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. 12 International Agency for Research on Cancer, World Health Organization, Lyon, France. ¹³European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI, Wellcome Genome Campus, Hinxton, UK. 14 Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. 15 Department of Medicine, University of Toronto, Toronto, Ontario, Canada. 16 Division of Medical Oncology and Hematology, University Health Network, Toronto, Ontario, Canada. 17 Department of Molecular Haematology, Norwich Medical School, The University of East Anglia, Norwich, UK. 18 Department of Haematology, Norfolk and Norwich University Hospitals NHS Trust, Norwich, UK. 19 Public Health Directorate, Asturias, Spain. 20 Hellenic Health Foundation, Athens, Greece. ²¹2nd Pulmonary Medicine Department, School of Medicine, National and Kapodistrian University of Athens, "ATTIKON" University Hospital, Haidari, Athens, Greece. ²²Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy. 23 Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria ibs.GRANADA, Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain. 24CIBER Epidemiology and Public Health CIBERESP, Madrid, Spain. 25Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain. 26 Navarra Public Health Institute, Pamplona, Spain. 27 Navarra Institute for Health Research, Pamplona, Spain. 28 Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. 29 Cancer Registry and Histopathology Department, Civic-M. P. Arezzo Hospital, Azienda Sanitaria Provinciale, Ragusa, Italy. 30 Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute - ISPO, Florence, Italy, 31 Department of Epidemiology, German Institute of Human Nutrition (DIfE), Potsdam-Rehbrücke, Germany, 32 Dipartimento Di Medicina Clinica E Chirurgia, Federico II University, Naples, Italy, 33 Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 34 Clinical Cooperation Unit Molecular Hematology/Oncology, German Cancer Research Center (DKFZ) and Department of Internal Medicine V, University of Heidelberg, Heidelberg, Germany. 35 Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy. 36 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. 37 Italian Institute for Genomic Medicine, Torino, Italy. 38 Unit of Nutrition and Cancer, Cancer Epidemiology Research Program and Translational Research Laboratory, Catalan Institute of Oncology, ICO-IDIBELL, Barcelona, Spain. 39University of Cambridge, Cambridge, UK. 40Division of Environmental Epidemiology and Veterinary Public Health, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands. 41 Department of Haematology, University of Cambridge, Cambridge, UK. 42 Center for Molecular Oncology and Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. 43 Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. 44 Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, UK. 45 Division of Hematology, Rambam Healthcare Campus, Haifa, Israel. 46 These authors contributed equally: Sagi Abelson, Grace Collord. 47These authors jointly supervised this work: Moritz Gerstung, John E. Dick, Paul Brennan, George S. Vassiliou, Liran I. Shlush. *e-mail: moritz. gerstung@ebi.ac.uk; John.Dick@uhnresearch.ca; BrennanP@iarc.fr; gsv20@sanger.ac.uk; Iiranshlush3@gmail.com

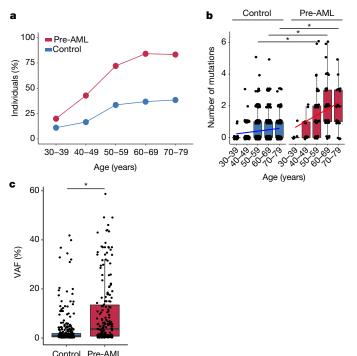


Fig. 1 | Prevalence of ARCH, number of mutations and clone size in individuals who developed AML. a, Prevalence of ARCH-PD among pre-AML cases (red) and controls (blue). b, The number of ARCH-PD mutations detected in cases and controls according to age. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5× interquartile range, respectively. Individual values are indicated as dots. c, VAF of ARCH-PD mutations. *P < 0.0005, two-sided Wilcoxon rank-sum test with Bonferroni multiple testing correction. All panels show data for n = 800 biologically independent samples.

a prevalence that is in line with the largest studies of ARCH in the general population⁴ ($P < 2.2 \times 10^{-16}$, two-sided Fisher's exact test; Extended Data Fig. 1).

The median number of ARCH-PD mutations per individual increased with age and was significantly higher in the pre-AML group relative to controls (Fig. 1b and Supplementary Table 2). Furthermore, examination of ARCH-PD VAF distribution revealed significantly larger clones among the pre-AML cases ($P = 1.2 \times 10^{-13}$, twosided Wilcoxon rank-sum test; Fig. 1c). To gain insight into clonal growth dynamics, we examined serially collected samples that were available for a subset of the validation cohort. We did not find significant differences in clonal expansion rates between pre-AML cases and controls (Extended Data Fig. 2a, b), although this may in part reflect the shorter follow-up of pre-AML cases, small sample size and large variance in growth rates (Extended Data Fig. 2c). The observed differences between pre-AML cases and controls may arise through cell-intrinsic or -extrinsic factors. Although these variables have not been adequately studied in ARCH, a number of observations in different contexts, such as aplasia, advanced age and after chemotherapy, have shown that increased clonal fitness is associated with distinct mutations depending on context^{10–12}. Notably, mutations in splicing factor genes were significantly enriched among the pre-AML cases relative to the controls (odds ratio, 17.5; 95% confidence interval, 8.1–40.4; $P = 5.2 \times 10^{-16}$ two-sided Fisher's exact test) and were present in significantly younger individuals (median age 60.3 compared to 77.3 years, $P = 1.7 \times 10^{-4}$, two-sided Wilcoxon rank-sum test; Fig. 2a). Previous work suggests that spliceosome mutations appear to confer a competitive advantage in the context of ageing¹⁰. Therefore, it is possible that the significantly higher prevalence of such clones in younger pre-AML cases may reflect extrinsic selection pressures rather than earlier mutation acquisition.

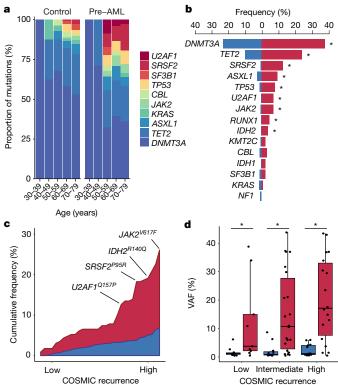


Fig. 2 | Accumulation of specific recurrent AML mutations in healthy individuals at a young age is associated with progression to AML. a, Relative frequency of mutations in the indicated genes according to age group for pre-AML cases and controls. b, Proportion of pre-AML cases (red) and controls (blue) who had ARCH-PD mutations in recurrently mutated genes. *P < 0.05, Fisher's exact test with Bonferroni multiple testing correction. c, The cumulative frequency of recurrent AML mutations (reported in >5 specimens in COSMIC) in pre-AML cases and controls. ARCH-PD mutations are ranked from left to right along the x axis from low to high recurrence. **d**, VAF of recurrent mutations in pre-AML cases and controls. Low, intermediate and highly recurrent COSMIC mutations are defined as those reported in 5-19 samples, 20-300 samples and >300 samples, respectively. Box plots indicate median, first and third quartiles and 1.5× interquartile range. *P < 0.05, two-sided Wilcoxon rank-sum test with Bonferroni multiple testing correction. All panels show data for n = 800 unique individuals.

In line with previous reports^{5,6}, we found that *DNMT3A* and *TET2* were the most commonly mutated genes in both groups (Fig. 2b). We could not identify any canonical NPM1 mutations nor any FLT3internal tandem duplication mutations, consistent with these arising late in leukaemogenesis 10,13. Recurrent CEBPA mutations, which are implicated in around 10% of de novo AML14, were also absent, suggesting that driver events in this gene may also be late events in AML evolution. In order to quantify the effect of different mutations on the likelihood of progression to AML, we ranked ARCH-PD mutations based on the number of times that they have been reported in Catalogue of Somatic Mutations in Cancer (COSMIC) database among individuals with haematological malignancies¹⁵. We found that mutations that are highly recurrent in cancer specimens were more common in pre-AML cases than in controls with ARCH-PD, whereas driver events in the controls tended to affect loci that are less frequently mutated in haematological malignancies and occurred at significantly lower VAF (Fig. 2c, d). Overall, these findings demonstrate notable differences in the mutational landscape of ARCH and pre-AML. Moreover, this work, in conjunction with recent insights into the origins of AML relapse¹⁶, suggests that AML progression typically occurs over many years through clonal evolution of preleukaemic HSPCs before acquisition of late mutations leads to overt malignant transformation.

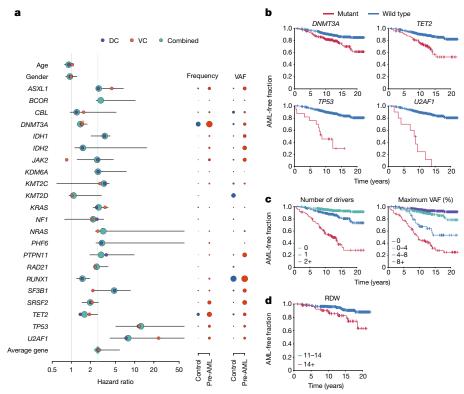


Fig. 3 | **Model of future risk of AML. a**, Forest plot of the risk of AML. Purple, orange and green circles indicate hazard ratios for the discovery (DC), validation (VC) and combined cohort, respectively. The horizontal lines denote 95% confidence intervals for the combined cohort. For each gene, the indicated hazard ratio applies to the 10-year risk of AML development conferred by each 5% increase in mutation VAF. The green vertical line indicates the mean hazard ratio across all genes. The hazard ratio for *RUNX1* must be interpreted with caution owing to the relatively high prevalence of deleterious germline variants in this gene, which may not be readily distinguishable from somatic mutations in unmatched

sequencing assays (see Methods). The proportion of individuals with mutations in each gene and the average VAF are indicated to the right of the forest plot; red and blue circles represent pre-AML cases and controls, respectively, with circle sizes scaled to reflect mutation frequency and VAF. \mathbf{b} - \mathbf{d} , Kaplan–Meier curves of AML-free survival, defined as the time between sample collection and AML diagnosis, death or last follow-up. Survival curves are stratified according to mutation status for selected genes (\mathbf{b}), number of driver mutations per individual and largest clone detected (\mathbf{c}) and RDW (\mathbf{d}). Data for n=796 unique individuals (\mathbf{a} - \mathbf{c}); n=299 individuals for whom RDW measurements were available (\mathbf{d}).

On the basis of these findings, we next developed an approach to quantify the relative contributions of driver mutations and clone sizes to the risk of progressing to AML. We tested different regularised logistic and Cox proportional hazards regression approaches, which achieved similar performance in both the discovery cohort (concordance $(C) = 0.77 \pm 0.03$) and the validation cohort $(C = 0.84 \pm 0.05)$; Extended Data Figs. 3, 4 and Supplementary Table 3). Models that were only trained on data from the discovery or validation cohort had similar coefficients (Fig. 3a). We therefore combined the datasets for a more accurate analysis of the contributions of mutations in individual genes to risk ($C = 0.77 \pm 0.05$; area under curve, 0.79; Supplementary Table 3). Quantitatively, we found that driver mutations in most genes conferred an approximately twofold increased risk of developing AML per 5% increase in clone size (Fig. 3a and Supplementary Table 3). Notable exceptions to this trend are the most frequently mutated ARCH genes, DNMT3A and TET2, which confer a lower risk of progression to AML (Fig. 3a, b and Supplementary Table 3). By contrast, a larger effect size was apparent for TP53 (hazard ratio, 12.5; 95% confidence interval, 5.0–160.5) and U2AF1 (hazard ratio, 7.9; 95% confidence interval, 4.1–192.2) mutations (Fig. 3a, b). However, we note that other ARCH-PD genes, such as SRSF2, can contribute a similar relative risk owing to their presence at a higher VAF in pre-AML cases (Fig. 3a, Extended Data Fig. 5a and Supplementary Note). Of note, mutations in TP53 and spliceosome genes (including U2AF1) are also associated with a poorer prognosis in AML¹⁴. Because the effect of each ARCH-PD mutation is deleterious and the effect of multiple mutations that are present in the same individual is multiplicative, a higher number of mutations is predicted to increase the risk of progression to AML (Fig. 3c). Similarly,

the size of the largest driver clone was also strongly associated with the risk of progression to AML, in agreement with the risk of individual mutations generally being proportional to VAF (Fig. 3c). Collectively, although the VAF and the number of mutations confer much of the predictive value, this model does demonstrate distinct gene-level risk factors, and is able to quantify the cumulative impact of multiple mutations and clonal size on the likelihood of progression to AML.

Although our predictive model performs well in identifying those at risk of developing AML in our experimental cohorts, AML incidence rates in the general population are low $(4:100,000)^1$, and thus millions of individuals would need to be screened to identify the few pre-AML cases, with many false positives. We therefore sought to determine whether routinely available clinical information could improve prediction accuracy or identify a high-risk population for targeted genetic screening. We first analysed complete blood count and biochemistry data that were available for 37 of the pre-AML cases and 262 controls. As reported previously^{5,10,17}, ARCH-PD was overwhelmingly associated with normal blood counts and this was also the case for pre-AML cases, indicating that these did not represent undiagnosed myelodysplastic syndrome¹⁸. We identified a significant association between higher red blood cell distribution width (RDW) and risk of progression to AML (P = 0.0016, Wald test with Bonferroni multiple-testing correction, Fig. 3d). Although traditionally used in the evaluation of anaemia, raised RDW has been correlated with inflammation, ineffective erythropoiesis, cardiovascular disease and adverse outcomes in several inflammatory and malignant conditions¹⁹. The correlation between RDW and risk of AML development remained highly significant when controls without ARCH-PD were excluded

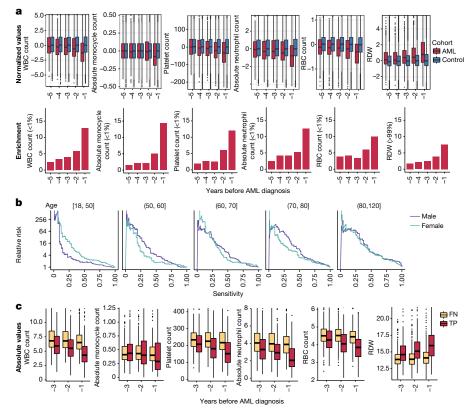


Fig. 4 | Increased risk of AML development inferred from electronic health records. a, Box plot of normalized laboratory measurements. Increased RDW, reduction in monocyte, platelet, red blood cell (RBC) and white blood cell (WBC) counts (top) show a high association (bottom) with a higher risk of AML development and differed at least a year before

AML diagnosis. **b**, Model performance stratification by age and gender. Age ranges are indicated above each graph. **c**, Absolute laboratory values for true positive (TP) and false negative (FN) predictions. Box plots indicate median, first and third quartiles and $1.5\times$ interquartile range.

from the analysis ($P = 3.5 \times 10^{-6}$, Wald test with Bonferroni multiple testing correction; Extended Data Fig. 5b). Higher RDW has previously been associated with ARCH and overall mortality⁵, but has never been shown to distinguish ARCH from pre-leukaemia. In order to verify RDW as a predictive factor and determine whether additional clinical parameters are associated with risk of AML development, we studied the Clalit database²⁰, which contains electronic health records that include an average of 3.45 million individuals per year and data that were collected over a 15-year period²¹. We identified 875 cases with AML using stringent criteria based on diagnostic codes and treatment records (Extended Data Fig. 6 and Supplementary Table 4). Analysis of RDW trends revealed significantly raised measurements several years before AML diagnosis relative to age and sex-matched controls (Fig. 4a). Additional parameters that correlated with risk of AML development included reductions in monocyte, platelet, red blood cell and white blood cell counts, albeit usually remaining above the thresholds for clinically relevant cytopenias 18 (Fig. 4a and Extended Data Fig. 7). These findings suggest that evolving de novo AML may sometimes have a considerable prodrome with subtle but discernible clinical manifestations. We next applied a machine-learning approach to construct an AML prediction model based entirely on variables that are routinely documented in electronic health records (Extended Data Fig. 8 and Supplementary Table 4). This model was able to predict AML 6-12 months before diagnosis with a sensitivity of 25.7% and overall specificity of 98.2%. The model performed consistently across different age groups with an increased relative risk of 28 and 24 for males and females, respectively, between the age of 60 and 70 years (Fig. 4b). To better understand which patients are most likely to be accurately classified by this model, we compared absolute laboratory values for true positives and false negatives. We found that 35.5% of false-negative predictions were for patients for whom infrequent blood count data were available (Extended Data Fig. 9). Some of the true-positive cases

had mildly abnormal blood counts that would not initiate a diagnostic work-up (Fig. 4c), and cytopenias that would be compatible with undiagnosed myelodysplastic syndrome¹⁸ were uncommon.

Collectively, our findings provide new insights into the pre-clinical evolution of AML and support the hypothesis that individuals at high risk of AML development can be identified years before they develop overt disease. To this end, we present two distinct models for the prediction of de novo AML: one based on somatic point mutations and the other on routinely documented clinical information. We find that basic clinical and laboratory data can identify a high-risk subgroup 6-12 months before AML presentation, while genetic information can identify a substantial fraction of cases several years to more than a decade before diagnosis. By characterizing features that distinguish benign ARCH from pre-leukaemia, our models give valuable insights into leukaemogenesis. It is evident from the current study, together with our recent analysis of mutation acquisition from pre-leukaemic development through to relapse¹⁶, that long-term pre-leukaemic HSPCs frequently carry mutations and undergo considerable clonal expansion while retaining differentiation capacity for years before AML diagnosis. Furthermore, it is clear that some mutations, particularly those affecting TP53 and U2AF1, impart a relatively high risk of subsequent AML, whereas mutations in other genes, for example *DNMT3A* and *TET2*, confer a lesser risk of malignant transformation. Previous studies suggest that oncogenic mutations in TP53 and spliceosome genes confer little or no competitive advantage in the absence of particular selective pressures^{11,22}, indicating that cell-extrinsic factors may be important determinants of clonal trajectory.

Cancer predictive models have enabled successful early detection and intervention programmes for several solid tumours^{23–25}. However, screening tests are unavailable for the sub-clinical stages of most haematological malignancies. Our study provides proof-of-concept for the feasibility of early detection of healthy individuals at high risk

of developing AML, and is a first step in the design of future clinical studies to investigate the potential benefits of early interventions in this deadly disease. However, the infrequency of AML necessitates that future screening tests provide high sensitivity and specificity. Our findings suggest that basic clinical data may identify a higher risk population that might benefit from targeted genetic screening. Equally, combining clinical and genetic information in a single model and including structural driver events is likely to improve model accuracy further. Nevertheless, establishing the utility of such a tandem approach will require extensive clinical and genetic analysis on the same population cohort, in a prospective setting. Furthermore, ARCH is associated with several non-malignant conditions 4,5, and may have a causal role in cardiovascular disease^{26,27}. Therefore, genetic testing for ARCH may also prove useful in the management of common age-related diseases. Moreover, this study has broader implications for cancer screening and early intervention beyond AML. Advances in sequencing technologies have revealed a remarkable degree of somatic genetic diversity in normal ageing tissues, often characterized by the presence of clones that have canonical oncogenic mutations²⁸. The degree to which clones at high risk of malignant transformation can be reliably distinguished from their indolent counterparts is an important biological question with compelling clinical ramifications. Understanding the selective pressures and cell-intrinsic mechanisms governing clonal fate is the next important step in developing strategies to predict and prevent progression to overt malignancy.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-

Received: 9 July 2017; Accepted: 3 May 2018; Published online 9 July 2018.

- Deschler, B. & Lübbert, M. Acute myeloid leukemia: epidemiology and etiology. Cancer 107, 2099-2107 (2006).
- Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc. Natl Acad. Sci. USA 111, 2548-2553
- Shlush, L. I. et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
- Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. 371, 2477-2487 (2014).
- Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. 371, 2488-2498 (2014).
- Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. 20, 1472–1478 (2014)
- Busque, L. et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood 88, 59-65 (1996).
- Shlush, L. I. Age-related clonal hematopoiesis. Blood 131, 496–504 (2018).
- Acuna-Hidalgo, R. et al. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am. J. Hum. Genet. **101**, 50–64 (2017).
- 10. McKerrell, T. et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep. 10, 1239-1245 (2015).
- 11. Wong, T. N., et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature 518, 552-555 (2015).
- Yoshizato, T. et al. Somatic mutations and clonal hematopoiesis in aplastic anemia. N. Engl. J. Med. 373, 35-47 (2015).
- Krönke, J. et al. Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. Blood 122, 100-108 (2013).
- 14. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. 374, 2209-2221 (2016).
- 15. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. **45**, D777–D783 (2017).
- Shlush, L. I. et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature 547, 104–108 (2017).
- 17. Buscarlet, M. et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood **130**, 753–762 (2017)
- 18. Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391–2405 (2016).
- Hu, L. et al. Prognostic value of RDW in cancers: a systematic review and meta-analysis. Oncotarget 8, 16027-16035 (2017).

- 20. Balicer, R. D. & Afek, A. Digital health nation: Israel's global big data innovation hub. Lancet 389, 2451-2453 (2017).
- 21. Dagan, N., Cohen-Stavi, C., Leventer-Roberts, M. & Balicer, R. D. External validation and comparison of three prediction tools for risk of osteoporotic fractures using data from population based electronic health records: retrospective cohort study. Br. Med. J. 356, i6755 (2017).
- McKerrell, T. & Vassiliou, G. S. Aging as a driver of leukemogenesis. Sci. Transl. Med. 7, 306fs38 (2015).
- Vickers, A. J. Prediction models in cancer care. CA Cancer J. Clin. 61, 315-326 (2011).
- Cassidy, A. et al. The LLP risk model: an individual risk prediction model for lung cancer. Br. J. Cancer 98, 270-276 (2008).
- Wang, X., Oldani, M. J., Zhao, X., Huang, X. & Qian, D. A review of cancer risk prediction models with genetic variants. Cancer Inform. 13, 19-28 (2014).
- Fuster, J. J. et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. Science 355, 842-847 (2017).
- Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular
- disease. *N. Engl. J. Med.* **377**, 111–121 (2017). 28. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. Science 349, 1483-1489 (2015).

Acknowledgements This work was supported by a Quest for Cure grant to L.I.S., J.C.Y.W. and M.D.M. from the Leukemia and Lymphoma Society, and the following grants to L.I.S from: ERC Horizon 2020 MAMLE, Abisch-Frenkel foundation and an American Society of Hematology Scholar Award. Further funding to J.E.D. was provided by the Canada Research Chair Program, Ontario Institute for Cancer Research, the province of Ontario, Canadian Cancer Society, the Canadian Institutes for Health Research and the Ontario Ministry of Health and Long Term Care to UHN, whose views are not expressed here. Work conducted at the Sanger Institute was supported by the Wellcome Trust and UK Medical Research Council. S.A. was personally funded by the Benjamin Pearl fellowship from the McEwen Centre for Regenerative Medicine, G.C. by a Wellcome Trust Clinical PhD Fellowship (WT098051); G.S.V. by a Wellcome Trust Senior Fellowship in Clinical Science (WT095663MA) and a Cancer Research UK Senior Cancer Research Fellowship (C22324/A23015). G.S.V.'s laboratory is also funded by the Kay Kendall Leukaemia Fund and Bloodwise. We thank A. Mitchell and all members of the Dick and Shlush laboratories for comments and T. Hudson for early study planning; G. Barabash for organising the Clalit dataset collaboration. The EPIC study centres were supported by the Hellenic Health Foundation, Regional Government of Asturias, the Regional Government of Murcia (no. 6236), the Spanish Ministry of Health network RTICCC (ISCIII RD12/0036/0018), FEDER funds/European Regional Development Fund (ERDF), "a way to build Europe" Generalitat de Catalunya, AGAUR 2014SGR726; EPIC Ragusa in Italy-Aire-Onlus Ragusa; Epic Italy-Associazione Italiana per la Ricerca sul Cancro (AIRC) Milan, Italy. S.V.B. and T.J.P. are supported by the Gattuso-Slaight Personalized Cancer Medicine Fund at the Princess Margaret Cancer Centre.

Reviewer information Nature thanks R. Levine, P. Van Loo and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.W.K.N., O.W., N.M.C. and E.N. contributed equally to the work. S.A. performed error-corrected sequencing, analysed sequencing data, performed statistical analyses, contributed to genetic predictive model derivation and wrote the manuscript. G.C. performed variant calling, statistical analyses, derived genetic predictive models and wrote the manuscript. M.G., S.W.K.N., O.W. and R.C. derived genetic predictive models. N.M.C., E.N. and N.B. derived the clinical prediction model. P.C.Z., Z.Z., I.C., K.N., C.L., C.H., D.H., F.M., J.E., J.K.M., D.P., L.T., P.K., S.V.B. and A.Br. and A.Ba. provided sequencing and technical support and enabled sample acquisition. L.H., Y.S., T.T.W., T.J.P., K.R. and D.J. provided bioinformatics support. R.L., S.H., M.J., K.M.B., A.Kr. and N.J.W. enabled sample acquisition, clinical data curation and/or provided clinical expertise. D.S., J.D.M., P.A., E.S., S.B., P.Be., M.D.M and I.M. contributed to data analysis and interpretation. P.J.C. and E.P. contributed to data interpretation and designed the targeted sequencing assay for the validation cohort. J.C.Y.W. revised the manuscript. J.R.Q., A.Ka., C.L.V., A.T., E.S.-F., J.M.H., R.C.T., R.T., G.M., H.B., S.Pa., R.K., S.S., S.Po., N.J.W., N.S., K.-T.K., M.F., J.M.K., E.R., P.V. and R.V. enabled sample acquisition (EPIC). A.T. and R.D.B. analysed Clalit data and derived the clinical prediction model. M.G. derived predictive genetic models, contributed to sequencing data analysis and manuscript writing. J.E.D. contributed to funding applications, study supervision and manuscript writing. P.Br. supervised sample acquisition from all EPIC centres. G.S.V. and L.I.S. designed and supervised all aspects of the study and wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0317-6.

Supplementary information is available for this paper at https://doi.org/ 10.1038/s41586-018-0317-6.

Reprints and permissions information is available at http://www.nature.com/ reprints.

Correspondence and requests for materials should be addressed to M.G., J.E.D., P.B., G.S.V. or L.I.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Study participants. Samples for both the discovery and validation cohort were obtained from participants in the EPIC study²⁹. All relevant ethical regulations were followed. Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki and protocols were approved by the relevant ethics committees (IARC Ethics Committee approval #14-31, the Weizmann Institute of Science Ethics board approval #60-1 and East of England–Cambridgeshire and Hertfordshire Research Ethics Committee reference number 98CN01). Patients with AML were identified based on the following ICD9 codes: 9861/3, 9860/3, 9801/3, 9866/3, 9891/3, 9867/3, 9874/3, 9840/3, 9872/3, 9895/3, 9873/3, which included only cases of de novo AML, and no secondary AML. All patients provided peripheral blood samples for which the buffy coat fractions were separated and aliquoted for long-term storage in liquid nitrogen before DNA extraction.

Discovery cohort. In total, 509 DNA samples were collected from individuals upon enrolment into the EPIC study between 1993 and 1998 across 17 different centres 29 (Supplementary Table 1). Altogether, 95 individuals who developed AML an average of 6.3 years (interquartile range (IQR) = 4.8 years) after the sample was collected were included in the pre-AML group. For the control group, 414 age- and gender-matched individuals were selected, as they did not develop any haematological disorders during the average follow-up period of 11.6 years (IQR = 2.1 years). The median age at recruitment was 56.7 years (range, 36.08–74.42). In order to minimize any possible demographic biases, an approximate 1:4.5 pre-AML to control ratio was maintained across the different centres.

Validation cohort. Samples were obtained from individuals enrolled in the EPIC-Norfolk longitudinal cohort study between 1994 and 2010. Samples and clinical metadata were available from 37 patients with AML (of which 8 were already included in the discovery cohort) and 262 age- and gender-matched controls without a history of cancer or any haematological conditions. The average time between the first blood sampling and AML diagnosis was 10.5 years (IQR = 8.3 years). The average follow-up period for the control cohort was 17.5 years (IQR = 3.8). For 12 individuals in the pre-AML cohort, 2–3 blood specimens were available, taken a median of 3.4 years apart. Of the 262 controls, 141 had multiple blood samples available, spanning a median of 10.5 years. Blood counts and other clinical parameters were available for all study participants (Supplementary Table 1).

Targeted sequencing. *Discovery cohort sequencing.* Targeted deep sequencing was performed using error-corrected sequencing as follows.

Shearing of genomic DNA, preparation of pre-capture sequencing libraries, hybridization-based enrichment, assessment of the libraries quality and enrichment following hybridization were performed as previously described 30 . In brief, 100 ng of genomic DNA was sheared before library construction (KAPA Hyper Prep Kit KK8504, Kapa Biosystems) with a Covaris E220 instrument using the recommended settings for 250-bp fragments. Following end repair and A-tailing, adaptor ligation was performed using 100-fold molar excess of Molecular Index Adaptor. Library clean-up was performed with Agencourt AMPure XP beads (Beckman-Coulter) and the ligated fragments were then amplified for eight cycles using 0.5 μ M Illumina universal and indexing primers.

Targeted capture was carried out on pools containing three indexed libraries. Each pool of adaptor-ligated DNA was combined with 5 μl of 1 mg ml⁻¹ Cot-I DNA (Invitrogen), and 1 nmol each of xGEN Universal Blocking Oligo, TS-p5, and xGen Universal Blocking Oligo, TS-p7 (8 nucleotides). The mixture was dried using a SpeedVac and then re-suspended in 1.1 μ l water, 8.5 μ l NimbleGen $2\times$ hybridization buffer and 3.4 μl NimbleGen hybridization component A. The mixture was heat denatured at 95 °C for 10 min before addition of 4 μl of xGen Lockdown Probes (xGen AML Cancer Panel v.1.0, 3 pmol). Each pool was then hybridized at 47 °C for 72 h. Washing and recovery of the captured DNA was performed according to the manufacturer's specifications. In brief, $100\,\mu l$ of clean streptavidin beads was added to each capture. Following separation and removal of the supernatant using a magnet, 200 μ l 1 \times Stringent Wash Buffer was added and the reaction was incubated at 65 $^{\circ}\text{C}$ for 5 min. The supernatant containing unbound DNA was removed before repeating the high stringency wash one additional time. Then, the bound DNA was washed as follows: (1) 200 μl 1× Wash Buffer I and separation of the supernatants by magnetic separation; (2) 200 μ l 1 \times Wash Buffer II after magnetic separation; (3) 200 μ l 1 \times Wash Buffer III and removal of the supernatants using magnetic separation. The captured DNA on beads was resuspended in 40 µl of Nuclease-Free water before dividing the total volume into two PCR tubes and subjecting the libraries to 10 cycles of post-capture amplification (manufacturer-recommended conditions; Kapa Biosystems). Before sequencing, libraries were spiked with 2% PhiX.

Validation cohort sequencing. Targeted sequencing was performed using a custom complementary RNA bait set (SureSelect, Agilent, ELID 0537771) designed

complementary to all coding exons of 111 genes that have been implicated in myeloid leukaemogenesis (Extended Data Table 1). Genomic DNA was extracted from peripheral whole blood and sheared using the Covaris M220. Equimolar pools of 10 libraries were prepared and sequenced on the Illumina HiSeq 2000 using 75-bp paired-end sequencing as per Illumina and Agilent SureSelect protocols.

Variant calling. *Discovery cohort variant calling and error correction.* The 126-bp paired-end reads sequencing data from the Illumina platform were converted to FASTQ format, the 2-bp molecular barcode information at each read of the pair was trimmed and was written in the reads' name. The thymine nucleotide required for ligation was removed from the sequences. Burrows-Wheeler aligner (BWA-mem)³¹ was used for alignment of the processed FASTQ files to the reference hg19 genome, after realignment of insertions and deletions (indels) using GATK³². An in-house algorithm was written to collapse read families that share the same molecular barcode sequence, the left-most genomic position of where each read of the pair maps to the reference and the CIGAR string. Families that consisted of at least two reads were used to generate consensus reads and a consensus base was called when there was at least 70% agreement. When a consensus base was called, it was assigned with the maximum base quality score observed in its corresponding pre-collapsed reads. Furthermore, when possible, duplex reads³³ were generated from two consensus reads, from a singleton read and a consensus read, or from two singleton reads. For each sequenced sample, we generated two BAM files, called BAM1 and BAM2. BAM1 consisted of duplex reads, consensus reads and singleton reads, thereby including some error-corrected and non-error corrected reads, while still containing all the genomic information encoded in the data in the form of unique DNA molecules. BAM2 consisted of duplex reads and consensus reads but not singleton reads. Both files were then analysed to detect single nucleotide variants (SNVs) and small indels using Varscan2³⁴. To further remove sequencing artefacts and improve sensitivity, we applied a two-step polishing statistical approach that models the error rate for each sequenced genomic position. For both steps, BAM1 was used and all samples except the sample that was investigated were included for error rate modelling. At step one, as previously described³⁰, the error rates were modelled by fitting Weibull distribution curves to the non-reference allele fractions. SNVs with allele fractions that were statistically distinguishable from the background error rates (P = 0) were further analysed. At step 2, the coverage of the non-reference allele fractions was considered using linear line fitting that describes the negative correlation that exist between the log(non-reference allele fraction) and the corresponding log(coverage) values. This allowed us to estimate different error rates at different coverage depths. Because indel errors are rare and cannot be appropriately modelled by the same statistical framework, they were called using barcode-mediated error correction alone. At least 10 consensus reads, 5 supporting reads on the forward strand, 5 supporting reads on the reverse strand and 2 duplex reads were required to call an indel. Additional post-processing steps applied to data from both the discovery cohort and validation cohort are detailed in 'Additional post-processing filters applied to discovery and validation cohort data'. Variants were annotated using Annovar³⁵. Validation cohort variant calling. Sequencing reads were aligned to the reference genome (GRCh37d5) using the Burrows-Wheeler aligner (BWA-aln)³¹. Unmapped reads, PCR duplicates and reads mapping to regions outside the target regions (merged exonic regions and 10 bp either side of each exon) were excluded from analysis. Sequencing depth at each base was assessed using Bedtools coverage v.2.24.0³⁶.

Somatic SNVs were called using shearwater, an algorithm developed for detecting subclonal mutations in deep-sequencing experiments (https://github.com/gerstung-lab/deepSNV v.1.21.5)^{37–39} considering only reads with minimum nucleotide and mapping quality of 25 and 40, respectively. This algorithm models the error rate at individual loci using information from multiple unrelated samples. Additionally, allele counts at the recurrent AML mutation hotspots listed in 'Curation of oncogenic variants' were generated using an in-house script (https://github.com/cancerit/alleleCount) and manually inspected in the Jbrowse genome browser⁴⁰. To further complement our SNV calling approach, we applied an extensively validated in-house version of CaVEMan v.1.11.2 (Cancer variants through expectation maximization)⁴¹. CaVEMan compares sequencing reads between study and nominated normal samples and uses a naive Bayesian model and expectation-maximization approach to calculate the probability of a somatic variant at each base (https://github.com/cancerit/CaVEMan).

Post-processing filters required that the following criteria were met for CaVEMan to call a somatic substitution. (1) If coverage of the mutant allele was less than 8, at least one mutant allele was detected in the first two-thirds of the read. (2) Less than 3% of the mutant alleles with base quality \geq 15 were found in the nominated normal sample. (3) Mean mapping quality of the mutant allele reads was \geq 21. (4) The mutation does not fall in a simple repeat or centromeric region. (5) Fewer than 10% of the reads covering the position contained an indel according to mapping. (6) Less than 80% of the reads report the mutant allele at the same read position. (7) At least a third of the reads calling the variant had a base quality

of 25 or higher. (8) Not all mutant alleles reported in the second half of the read. (9) Position does not fall within a germline insertion or deletion.

The following additional post-processing criteria were applied to all SNV calls. (1) Minimum VAF = 0.5% with a minimum of five bidirectional calls reporting the mutant allele (with at least two reads in forward and reverse directions). (2) No indel called within a read length (75 bp) of the putative substitution.

Small indels were sought using two complementary bioinformatics approaches. First, an in-house version of Pindel v.2. 2^{42} (https://github.com/cancerit/cgpPindel) was applied. We additionally used the aforementioned deepSNV algorithm in order to increase sensitivity for indels present at low VAF. VAF correction was performed using an in-house script (https://github.com/cancerit/vafCorrect).

Post-processing filters required that the following criteria were met for a variant to be called. (1) A minimum of five reads supporting the variant with a minimum of two reads in each direction. For Pindel, the total read count was based on the union of the BWA and Pindel reads reporting the mutant allele. (2) VAF \geq 0.5%. (3) Variant not present within an unmatched normal panel of approximately 400 samples. (4) No reads supporting the variant identified in the nominated normal sample.

Mutations were annotated according to ENSEMBL v.58 using VAGrENT⁴³ for transcript and protein effects (https://github.com/cancerit/VAGrENT) and Annovar³⁵ for additional functional annotation.

Additional post-processing filters applied to discovery and validation cohort data. The following variants were flagged for additional inspection for potential artefacts, germline contamination or index-jumping event. (1) Any mutant allele reported within 75 bp of another variant. (2) Any mutant allele with a population allele frequency >1 in 1,000 according to any of five large polymorphism databases (ExAC, 1000 Genomes Project, ESP6500, CG46 and Kaviar) that is not a canonical hotspot driver mutation with COSMIC recurrence >100. (3) Mutations that were present in >10% of the control cohort but not recurrent in COSMIC were flagged as potential germline variants or sequencing artefacts. (4) As artefactual indels tend to be recurrent, any indels occurring in >2 samples were flagged as for additional inspection.

Curation of oncogenic variants. Putative oncogenic variants were identified according to evidence for functional relevance in AML as previously described and used to define ARCH-PD¹⁴.

Variants were annotated as likely driver events if they fulfilled any of the following criteria. (1) Truncating mutations (nonsense, essential splice site or frameshift indel) in the following genes implicated in AML pathogenesis by loss-of-function: NF1, DNMT3A, TET2, IKZF1, RAD21, WT1, KMT2D, SH2B3, TP53, CEBPA, ASXL1, RUNX1, BCOR, KDM6A, STAG2, PHF6 and KMT2C. (2) Truncating variants in CALR exon 9. (3) JAK2^{V617F}. (4) FLT3 internal tandem duplication. (5) Nonsynonymous variants at the following hotspot residues: CBL E366, L380, C384, C404, R420 and C396; DNMT3A R882; FLT3 D835; IDH1 R132; IDH2 R172 and R140; KIT W557, V559 and D816; KRAS A146, Q61, G13 and G12; MPL W515; NRAS Q61, G12 and G13; SF3B1 K700 and K666; SRSF2 P95; U2AF1 Q157, R156 and S34. (6) Non-synonymous variants reported at least 10 times in COSMIC with VAF <42% and population allele frequency <0.003. (7) Non-synonymous variants clustering within a functionally validated locus or within four amino acids of a hotspot variant with population allele frequency < 0.003 and VAF < 42%. (8) Non-synonymous variants reported in COSMIC > 100 times with population allele frequency < 0.003 regardless of VAF.

Our driver curation strategy inevitably runs a small risk of including germline variants in familial AML genes. We feel that in the real world, where a matched constitutional DNA sample would be unavailable, this is the best approach.

Statistical analysis. All statistical analyses were performed in the R statistical programming environment. A two-sided Wilcoxon rank-sum test was used to assign significance level for differences in the median number of somatic mutations among the pre-AML and control groups, the median VAF of mutations among groups. and the age of individuals with spliceosome mutations. Fisher's exact test was used to assess the significance of differences in the prevalence of ARCH among the groups and spliceosome mutations in the pre-AML group.

Predictive modelling. Cox proportional hazards model with random effects. We used a Cox proportional hazards regression to model AML progression-free survival as previously described ^{14,38}. We used random effects for the Cox proportional hazards model in the CoxHD R package (http://github.com/gerstung-lab/CoxHD). A key strength of this approach is the ability to include many variables in one model while shrinking estimated effects for parameters with weak support in the data, thus controlling for overfitting. We used weighting to minimize the biases introduced by the artificial case–control ratio ^{44,45} and calculated hazard ratios relative to the (approximate) true cumulative incidence of about 1–3/1,000 in the given age range over a follow up of 10–20 years. The observed driver mutation frequency and VAF in pre-AML cases closely resembled values expected based on the estimated risks, indicating that risk model and driver prevalence are well aligned (Extended Data Fig. 4). Full details of model derivation and comparisons

with alternative methods are included in the accompanying code (Supplementary Note, also available at https://github.com/gerstung-lab/preAML). In brief, variables comprised age, gender and the VAF of putative driver mutations (see 'Curation of oncogenic variants' for details of variant curation). We performed agnostic imputation of missing variables by mean and linear rescaling of gene variables by a power of 10 to a magnitude of 1. The model was first trained separately on the discovery cohort and validation cohort. For each of these two models, we evaluated the following measures of predictive accuracy before and after leave-one-out cross-validation (LOOCV): concordance $(C)^{46}$ and time-dependent area under the receiver-operating characteristic curve (AUC)⁴⁷. The models trained on the validation and discovery cohorts were then cross-validated using the data from the other cohort. In view of the cross-validation results and close correlation between coefficients (Supplementary Table 3), we derived a model on the combined cohorts using both cohorts in order to achieve greater accuracy on the individual effects. Confidence intervals were calculated using 100 bootstrap samples. The coefficients and performance metrics for each iteration of the model are included in Supplementary Table 3.

Concordance measures were obtained using the survConcordance() function implemented in the survival R package⁴⁵. Dynamic AUC was calculated with AUC.uno() implemented in the survAUC package. Time-independent AUCs were calculated using the performance function implemented in the ROCR package. The expected incidence of AML was calculated from the UK office of national statistics, available at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml/incidence. All-cause mortality data was obtained from the office of national statistics (https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables).

Ridge-regularized logistic regression. Using the same covariates as in 'Cox proportional hazards model with random effects', we fitted a ridge-regularized logistic regression model to dichotomised outcome data. While logistic regression is a common choice for case-control analyses, a downside of this approach is the inability to explicitly use time-dependent covariates. The penalty parameter was chosen using LOOCV on the full cohort; this value was then used on the discovery cohort and validation cohort to yield the same scaling of coefficients. Confidence intervals were calculated using 100 bootstrap samples. Fitting was performed using the glmnet R package. AUC as the primary performance metric was calculated using the ROCR R package.

Additional regression models. Two alternative predictive models were developed. Model 1 performs logistic-regression-based predictions using four types of features: gender, age at blood sampling, the sum of the VAFs ARCH-PD reported in COSMIC v.80 to be recurrent (at least two case reports in haematopoietic and lymphoid tissues) and somatic mutation burden of selected genes, where each gene was represented by the sum of the VAFs corresponding to ARCH-PD mutations in that gene. We measured the predictive performance of each gene via the AUC obtained in a fivefold cross-validation when using only the gene as a predictive feature, and only retained genes with AUC > 55% in the final model.

For model 2 we applied LASSO regression as implemented in the glmnet R package, while enabling LOOCV to fit a Cox regression model. A minimal subset of ARCH-PD variants was selected for which the respective weighted combined VAFs were highly predictive of AML development in the training set. Scores were calculated for each patient as a linear combination of VAF of mutations weighted by regression coefficients that were estimated from the training data. As most scores were zero in the training subset, non-zero scores were discretized to take on a value of 1 that corresponds to AML prediction.

Models 1 and 2 were trained on the discovery cohort and tested for their association with AML development using the validation cohort data. Survival analysis was performed using the Kaplan–Meier and Cox proportional hazards models. Wald's test was used to evaluate the significance of hazard ratios. Logistic regression models were used with the positive predictive value metric to determine the ability of various mutations and other patient parameters to predict AML development. The rms R package was used for logistic regression analysis, and the pROC 1.8 R package was used for receiver-operating characteristic curve analysis.

AML-predictive model based on electronic health records. Clalit database. The Clalit database includes information from patients covered by the Clalit health services in Israel²⁰ during the years 2002–2017. The Clalit training-set data, contains the electronic health records (EHR) of 3.45 million individuals per year on average. All data was anonymized through hashing of personal identifiers and addresses and randomization of dates by sampling a random number of weeks for each patient and adding it to all dates in the patient diagnoses, laboratory and medication records. This approach maintained differential data analysis per patient. Diagnoses codes were acquired from both primary care and hospitalization records, and were mapped to the ICD-9 coding system for historical reasons, with few exceptions that used a partial ICD-10 coding system. Laboratory records were normalized for age and gender by subtracting raw test values from the median

levels observed among all test values with matching gender and age (using a bin size of five years). We observed some chronological biases in laboratory ranges, but avoid normalizing these and instead insured case and controls are matched for chronological distributions.

Defining AML cases. We screened for all active patients (18 < age < 100) who were diagnosed with AML (ICD-9 code 205.0*) between the years 2003 and 2016. We then excluded cases based on the following criteria. (1) We excluded patients with prior myeloid malignancies to omit secondary AML, consistent with the case selection for the genetic model. The following diagnosis were excluded if documented within five years before the diagnosis of AML: essential thrombocythemia (ICD-9 238.71), low-grade myelodysplastic syndrome (MDS) (ICD-9 238.72); high-grade MDS lesions (ICD-9 238.73); MDS with 5q deletion (ICD-9 238.74); MDS, unspecified (ICD-9 238.75); polycythemia vera (ICD-9 238.4); myelofibrosis (ICD-9 289.83); chronic myelomonocytic leukaemia (ICD-9 206.10-206.22).

- (2) Patients that had any procedures performed on bone marrow or spleen (ICD-10 code Z41) in the five-year period before first mention of AML diagnosis code in their record. These patients were presumed to have an inaccurate AML diagnosis date or misdiagnosis recorded.
- (3) Patients that received medications suggestive of an alternative diagnosis of chronic myeloid leukaemia, lymphoid malignancy or acute promyelocytic leukaemia (APL). At any time before diagnosis: imatinib, dasatinib, anagrelide, hydroxycarbamide, asparaginase, pegaspargase or arsenic trioxide. At any time after diagnosis: imatinib, dasatinib, methotrexate, tretinoin or arsenic trioxide. At any time after diagnosis, along with any acute lymphoblastic leukaemia diagnosis (ICD-9 204) or more than single dose: mercaptopurine. APL cases were excluded as early diagnosis of APL will most probably not change its outcome, as treatment is successful already.
- (4) Patients without a hospitalization record within three months before or after the onset diagnosis. This parameter was used as it is unlikely that a patient with AML will not be hospitalized close to diagnosis. This filter reduced false-positive cases and better defined the onset date.

We refined the estimated time of onset using the earliest time at which any of the following diagnosis appeared in the patient's history: amyloidosis (ICD-9 277.3), lymphoid leukaemia (ICD-9 204), myeloid leukaemia (ICD-9 205), leukaemia of unspecified cell type (ICD-9 208).

This strategy retained 875 AML cases in the training set for further analysis. These were further validated by manual expert inspection of the complete records of 8% of the cases.

To define the control set, we included all Clalit individuals that were not cases. Since our analysis was aggregating data from a historical time window of 15 years, we associated each control with a randomized time point for evaluation. Using this approach, both cases and controls represented a specific time point in the historical record of a patient, with matching calendric, age and gender distributions. Through this strategy 5,238,528 controls were used.

Defining features for construction of a predictive a score. We extracted the following features for discriminative analysis of cases and controls (this procedure was applied repeatedly in cross-validation as discussed below). (1) Age (in years) at time point. (2) Gender. (3) Laboratory features. Out of 2,770 different types of laboratory tests, we selected the top 50 most frequent laboratory tests (Supplementary Table 4). For each laboratory measurement, we used median age- and gendernormalized test values per patient in three time windows for 6-12 months before onset, 1-2 years before onset and 2-3 years before onset. In addition, we compute the slope of the normalized laboratory measurements for the 6-12 month time window using a linear regression model. (4) Diagnosis features. Of the 1780 different major ICD-9 diagnosis codes, we selected only diagnoses that were previously observed in at least 10 different cases and have an increased relative risk for AML >twofold (as observed in the training set, Supplementary Table 4). For each diagnosis code, we mark whether it appeared in each of the patients in time intervals of 6 months to 3 years, and 3-5 years before onset. (5) BMI features. For each patient in the cohort, we extracted median BMI, weight and height as measured in time intervals of 6 months to 2 years, and 2-3 years before onset.

Gradient boosting. We used the R package xgboost to infer parameters for a classifier given cases and controls. Objective was set to binary:logistic, the evaluation metric to AUC. We set nrounds = 5000, eta = 0.001, gamma = 0.1, lambda = 0.01, alpha = 0.01, max_depth = 6, min_child_weight = 2, subsample = 0.7 and colsample_bytree = 0.7. The boosting algorithm reports a function f that computes a predictive score given the features. Given a threshold T the expression f(patient features) > T defines a classifier. To standardise thresholds we estimate quantiles for the scores on the training set T(p) = quantile(f(train),p) and define the classifier for specificity level p as f(patient features) > T(p) (Supplementary Table 4). Cross-validation and relative risk evaluation. To evaluate the predictive value of the classification scheme while considering the strong age and gender biases in the incidence of AML, we performed fivefold cross-validation after splitting the

cases and controls into five age- and gender-matched groups. For each fold, we sampled 100,000 controls and combined with the cases, constructed the feature set and trained the model. The model was then tested on the fold cases along with 200,000 sampled controls. We used standardized classifier parameters and standardized thresholds that were inferred based on each training set to generate a series of classifications on each test set and merged these based on the control quantiles in the test as described above. Given a threshold p to define high and low prediction score, we counted for each bin b that defines a patient in a specific age (<40, 40-50, 50-60, 60-70, 70-80, >80) and gender group: the number of cases in bin b (N^b_{case}) and the number of controls in bin b ($N^b_{control}$) where N^b is the number of patients in bin b (entire database minus recall controls that are only a sample of the cohort). N^b (case, high score) = N^b_{TP} indicates the number of true positives (TP); N^b (case, low score) = N^b_{FP} indicates the number of false negatives (FN); N^b (control, high score) = N^b_{TP} indicates the number of false positives (FP); N^b (control, low score) = N^b_{TP} indicates number of true negatives (TN).

For each age and gender group, the absolute risk for AML in the bin is computed by $r^b_{\rm abs}=N^b_{\rm case}/N^b$. The absolute risk given a high score is estimated as $r^b_{\rm abs,high}=N^b_{\rm TP}/(N^b_{\rm FP}+N^b_{\rm TP})$. The relative risk in the bin is defined by ${\rm rr}^b=r^b_{\rm abs,high}/r^b_{\rm abs}$ where the sensitivity level for the classifier threshold level is defined as ${\rm sense}^b=N^b_{\rm TP}/N^b_{\rm case}$.

$$rr = \frac{\frac{\frac{TP \times cases}{(TP + FN)}}{\frac{TP \times cases}{(TP + FN)} + \frac{FP \times controls}{(FP + TN)}}}{\frac{cases}{cases + controls}}$$

Clonal growth rate calculation. Individual clones were defined by different mutations in different study participants. Per clone we calculated α according to the following equation:

$$a = \log(V/V_0) / (T - T_0)$$

where T and T_0 indicate the age of the individual at the two measurement time points. V and V_0 correspond to the VAF at T and T_0 , respectively.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

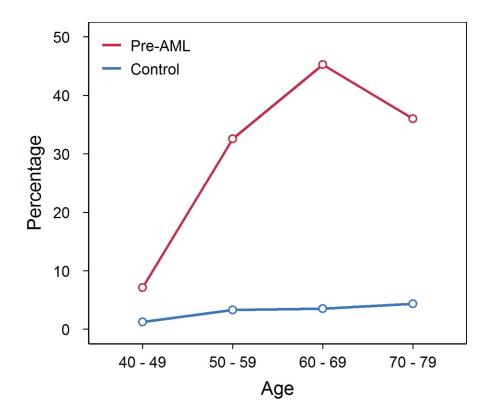
Code availability. Code for derivation of the prediction model is publically available on Github (https://github.com/gerstung-lab/preAML). Code for the analysis of error-corrected sequencing is available from the Shlush lab upon request.

Data availability. Targeted sequencing data for the discovery cohort are deposited as BAM files at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession number EGAD00001003583. All other data are available from the corresponding authors upon reasonable request. Sequencing data for the validation cohort are deposited at the European Genome-phenome Archive with accession number EGAD00001003703.

- Riboli, E. et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* 5, 1113–1124 (2002).
- Newman, A. M. et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat. Med. 20, 548–554 (2014).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595 (2010).
- McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).
- Kennedy, S. R. et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat. Protoc. 9, 2586–2606 (2014).
- Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576 (2012)
- Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 10, 1556–1566 (2015).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010).
- Gerstung, M. et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. Nat. Commun. 3, 811 (2012).
- Gerstung, M. et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat. Genet. 49, 332–340 (2017).
- Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348, 880–886 (2015).
- Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 17, 66 (2016).
- Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486, 400–404 (2012).
- Raine, K. M. et al. cgpPindel: identifying somatically acquired insertion and deletion events from paired end sequencing. *Curr. Protoc. Bioinformatics* 52, 15.17.1–15.7.12 (2015).
- Menzies, A. et al. VAGrENT: Variation Annotation Generator. Curr. Protoc. Bioinformatics 52, 15.18.1–15.18.11 (2015).

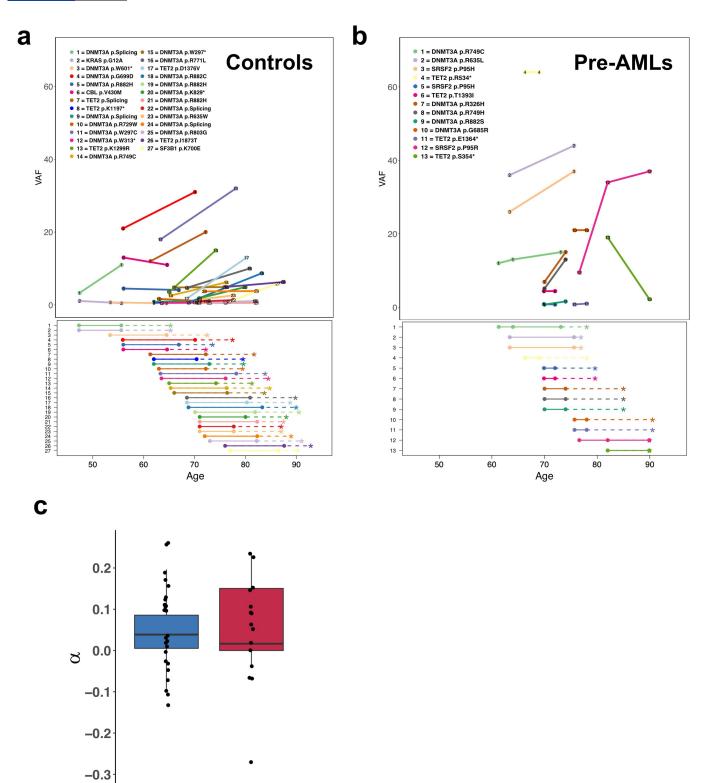


- 44. Antoniou, A. C. et al. A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. *Genet. Epidemiol.* 29, 1–11 (2005).
 45. Therneau, T. & Grambsch P. M. *Modeling Survival Data: Extending the Cox Model* 1st edn (Springer-Verlag, New York, 2000).
- $46. \ \ Harrell, F.\,E.\,Jr, Lee, K.\,L.\,\&\,Mark, D.\,B.\,Multivariable\ prognostic\ models: issues\ in$ developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387 (1996).
- 47. O'Quigley, J., Xu, R. & Stare, J. Explained randomness in proportional hazards models. Stat. Med. 24, 479–489 (2005).



No. with mutation				
Pre-AML	1	14	19	7
Control	1	8	9	9
Total				
Pre-AML	14	43	42	25
Control	82	242	254	161

Extended Data Fig. 1 | **Prevalence of ARCH-PD mutations with VAF** \geq **10% according to age.** Red and blue lines represent the proportion of pre-AML cases and controls, respectively, that had ARCH-PD mutations with VAF \geq 10%.

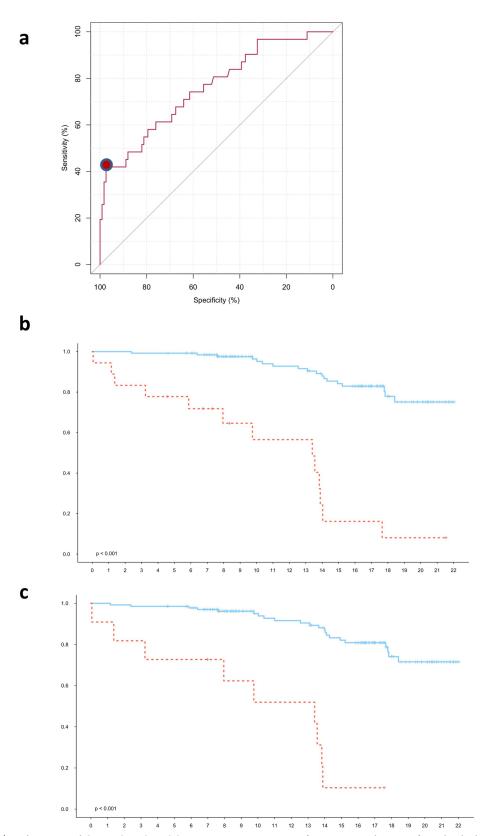


Extended Data Fig. 2 | Serially collected sampling supports a long-lived HSPCs as the cell of origin for most ARCH-PD clones. a, b, VAF trajectory of persistent clones carrying putative driver mutations in controls (a) and pre-AML cases (b). Age is indicated on the x axis. Top, VAF is shown on the y axis and each persistent mutation is shown in a different colour, with circles denoting individual serial samples and solid lines representing the growth trajectory between serial samples. Bottom,

Control

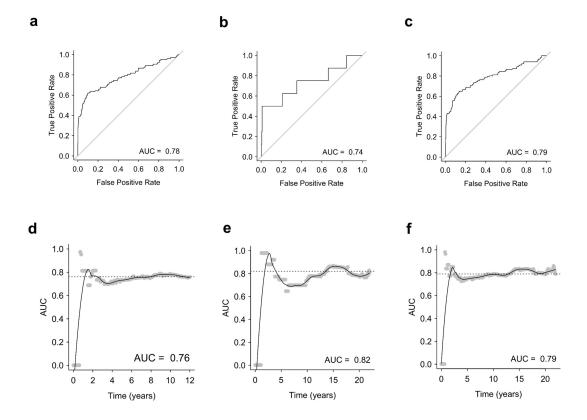
pre-AML

dashed lines indicate the time interval between the last sampling and the end of follow-up (controls) or AML diagnosis (cases). **c**, Clonal growth rates (α) are shown for 27 control clones corresponding to 54 time points and 13 pre-AML clones corresponding to 15 time points. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5 × interquartile range, respectively.

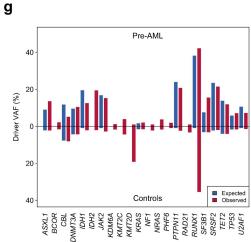


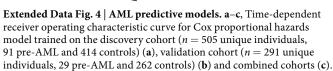
Extended Data Fig. 3 | Performance of the combined model in predicting progression to AML. a, Receiver operating characteristic curve for prediction of AML development using model 1 (see Methods). The red dot indicates the point on the curve with the highest positive predictive value with sensitivity of 41.9% and specificity of 95.7%. b, c, Kaplan–Meier

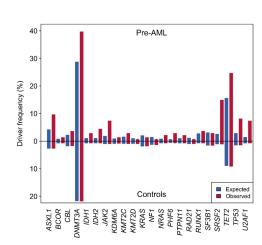
estimates of time to AML diagnosis for individuals predicted to develop AML (red) and not develop AML (blue) using model 1 (b; hazard ratio, $10.38; P=4.2\times10^{-10},$ Wald test) and model 2 (c; hazard ratio, $10.75; P=1.75\times10^{-8},$ Wald test), from the point of enrolment until the end of follow-up for patients enrolled in the EPIC study.



h

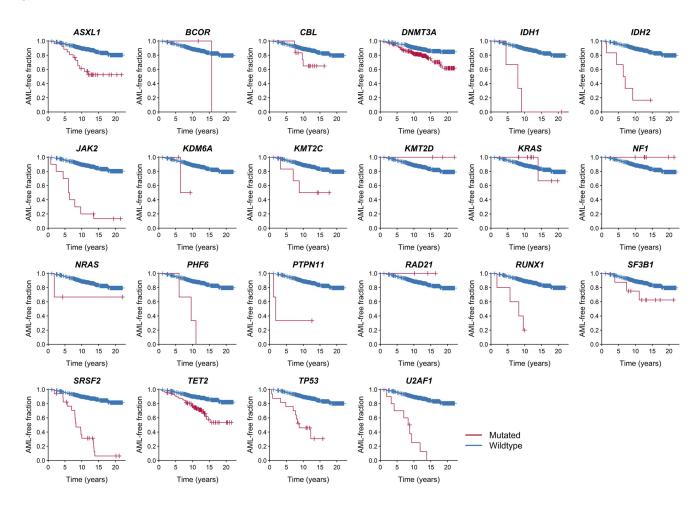


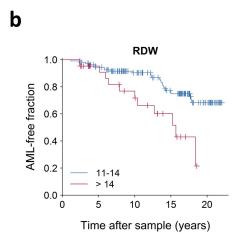




d–**f**, Dynamic AUC for Cox proportional hazards models trained on the discovery cohort (**d**), validation cohort (**e**) or combined cohort (**f**). **g**, **h**, Red and blue bars indicate the observed and expected VAF (**g**) and driver frequency (**h**) of pre-AML cases and controls for each gene indicated on the x axis.

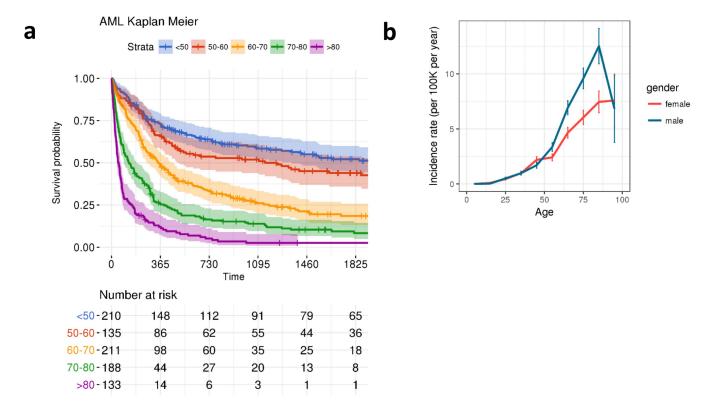
a





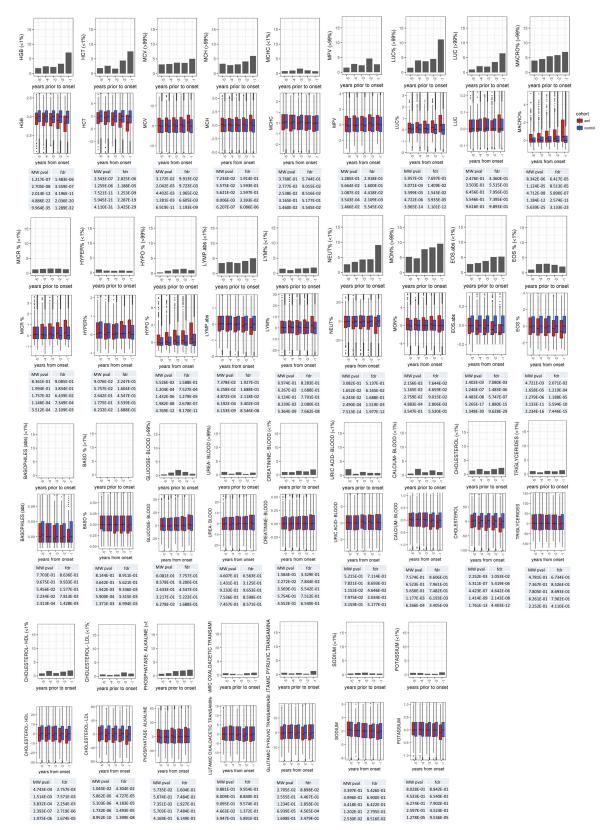
Extended Data Fig. 5 | AML-free survival based on mutation status and RDW. a, Kaplan–Meier curves of AML-free survival, defined as the time between sample collection and AML diagnosis, death or last follow-up. Survival curves are stratified according to mutation status in genes mutated in at least three samples across the combined validation and discovery cohorts. n=796 unique individuals. b, Kaplan–Meier curve of

AML-free survival stratified according to RDW value >14 or \leq 14. Plot represents data for n=128 biologically independent individuals who had RDW measurements, including all pre-AML cases regardless of ARCH-PD status, and controls with ARCH-PD (controls without detectable mutations were omitted).



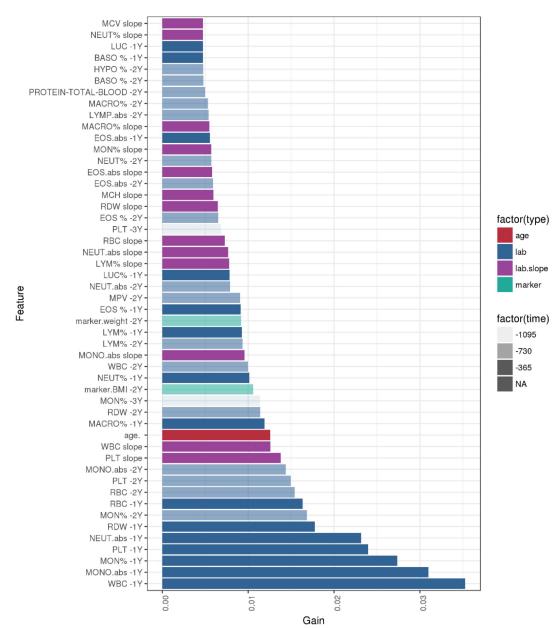
Extended Data Fig. 6 | **Description of the cohort and the EHR-derived measurements. a**, Kaplan–Meier curves showing age stratified survival rates for 875 individuals who developed AML. **b**, Line plot representation

of the number of cases per 100,000 control individuals in the EHR database. The centre values and error bars define the mean and s.d., respectively.



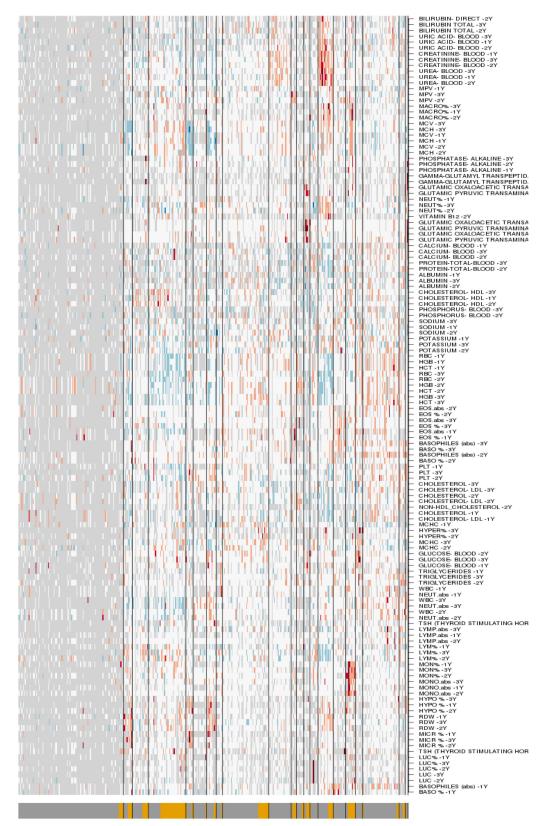
Extended Data Fig. 7 | Laboratory measurements contributing to the EHR model. Normalized laboratory measurements for pre-AMLs (red) and controls (blue) (middle) and their association (bottom) with higher risk of AML are shown. The grey bars indicate the percentage of pre-AML

cases with laboratory results either below the 1st percentile or above the 99th percentile. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5 \times interquartile range, respectively.



Extended Data Fig. 8 | Top 50 parameters for the EHR model. The relative contribution of the top 50 features incorporated into the EHR prediction model, ranked according to their predictive value (gain). 1Y, one year before AML diagnosis; 2Y, two years before AML diagnosis; 3Y, three years before AML diagnosis; BASO%, percentage of basophils; BMI, body mass index; EOS.abs, absolute eosinophil count; EOS%, percentage of eosinophils; HYPO%, percentage of hypochromia; LUC,

large unstained cells; LYM%, percentage of lymphocytes; LYMPH.abs, absolute lymphocyte count; MACRO%, percentage of macrocytosis; MCH, mean corpuscular haemoglobin; MCV, mean corpuscular volume; MON%, percentage of monocytes; MONO.abs, absolute monocyte count; MPV, mean platelet volume; NEUT.abs, absolute neutrophil count; NEUT%, percentage of neutrophils; PLT, platelet count; RBC, red blood cell count; RDW, red cell distributiom width; WBC, white blood cell count.



Extended Data Fig. 9 | See next page for caption.



Extended Data Fig. 9 | Distribution of EHR model parameters. Heat map illustrating absolute values of clinical measurements. Blue, white and red indicate low, intermediate and high values, respectively. Light grey indicates missing data. False-negative and true-positive annotations are indicated at the bottom as dark-grey and yellow colour bars, respectively. 1Y, one year before AML diagnosis; 2Y, two years before AML diagnosis; 3Y, three years before AML diagnosis; BASO%, percentage of basophils; EOS. abs, absolute eosinophil count; HCT, haematocrit; HDL; high density lipoprotein; HGB, haemoglobin; Hyper%, percentage of hyperchromia; Hypo%, percentage of hypochromia; LDL, low density lipoprotein; LUC, large unstained cells; LYM%, percentage

of lymphocytes; LYMP.abs, absolute lymphocyte count; MACRO%, percentage of macrocytosis; MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; MICR%, percentage of microcytosis; MON%, percentage of monocytes; MONO.abs, absolute monocyte count; MPV, mean platelet volume; PLT, platelet count; NEUT%, percentage of neutrophils; NEUT.abs, absolute neutrophil count; RBC, red blood cell count; RDW, red cell distribution width; Transamina, transaminase; Transpeptid., transpeptidase; TSH, thyroid stimulating hormone; WBC, white blood cell count.

Extended Data Table 1 \mid Genes sequenced by cRNA bait pull-down in the validation cohort

GNB1	FBXW7	CUL2	RAD51	CBLC
CSF3R	IRF1	CDH23	IDH2	U2AF2
MPL	CSF1R	PTEN	CREBBP	ASXL1
NRAS	NPM1	SMC3	SMG1	PTPRT
NOTCH2	PHACTR1	HRAS	CBFB	GNAS
RIT1	DAXX	WT1	CTCF	RUNX1
CACNA1E	PHIP	SF1	SMPD3	U2AF1
FAM5C	MYB	EED	PRPF8	CSF2RB
DNMT3A	FNDC1	CNTN5	TP53	CBX7
ASXL2	CUX1	MLL	NF1	EP300
SF3B1	MLL5	CBL	SUZ12	ZRSR2
IDH1	LUC7L2	ETV6	STAT5B	BCOR
CUL3	BRAF	KRAS	KANSL1	KDM6A
GIGYF2	CUL1	MLL2	DCAF7	GATA1
CBLB	EZH2	PRPF40B	SRSF2	SMC1A
GATA2	MLL3	PPFIA2	ASXL3	PHF8
STAG1	RAD21	SH2B3	SETBP1	MED12
PIK3CA	MYC	PTPN11	DNMT1	ATRX
FRYL	JAK2	FLT3	EPOR	RPS6KA6
KIT	CDKN2A	PDS5B	JAK3	DIAPH2
UGT2A3	HNRNPK	DCLK1	CEBPA	STAG2
TET2	NOTCH1	RB1	ZFP36	PHF6



Reprogramming human T cell function and specificity with non-viral genome targeting

Theodore L. Roth^{1,2,3,4,5}, Cristina Puig-Saus⁶, Ruby Yu^{3,4,5}, Eric Shifrut^{3,4,5}, Julia Carnevale⁷, P. Jonathan Li^{3,4,5}, Joseph Hiatt^{1,2,3,4,5}, Justin Saco⁶, Paige Krystofinski⁶, Han Li^{8,9}, Victoria Tobin^{3,4,5}, David N. Nguyen^{3,4,5}, Michael R. Lee⁴, Amy L. Putnam⁴, Andrea L. Ferris¹⁰, Jeff W. Chen¹¹, Jean-Nicolas Schickel¹¹, Laurence Pellerin^{12,13}, David Carmody¹⁴, Gorka Alkorta-Aranburu¹⁵, Daniela del Gaudio¹⁵, Hiroyuki Matsumoto¹⁶, Montse Morell¹⁶, Ying Mao¹⁶, Min Cho¹⁷, Rolen M. Quadros¹⁸, Channabasavaiah B. Gurumurthy¹⁸, Baz Smith¹⁶, Michael Haugwitz¹⁶, Stephen H. Hughes^{10,11}, Jonathan S. Weissman^{8,9}, Kathrin Schumann^{3,4,5}, Jonathan H. Esensten¹⁹, Andrew P. May¹⁷, Alan Ashworth⁷, Gary M. Kupfer²⁰, Siri Atma W. Greeley¹⁴, Rosa Bacchetta^{12,13}, Eric Meffre¹¹, Maria Grazia Roncarolo^{12,13}, Neil Romberg^{21,22}, Kevan C. Herold²³, Antoni Ribas^{6,24,25,26}, Manuel D. Leonetti^{8,9,28} & Alexander Marson^{3,4,5,7,17,27}*

Decades of work have aimed to genetically reprogram T cells for therapeutic purposes^{1,2} using recombinant viral vectors, which do not target transgenes to specific genomic sites^{3,4}. The need for viral vectors has slowed down research and clinical use as their manufacturing and testing is lengthy and expensive. Genome editing brought the promise of specific and efficient insertion of large transgenes into target cells using homology-directed repair^{5,6}. Here we developed a CRISPR-Cas9 genome-targeting system that does not require viral vectors, allowing rapid and efficient insertion of large DNA sequences (greater than one kilobase) at specific sites in the genomes of primary human T cells, while preserving cell viability and function. This permits individual or multiplexed modification of endogenous genes. First, we applied this strategy to correct a pathogenic IL2RA mutation in cells from patients with monogenic autoimmune disease, and demonstrate improved signalling function. Second, we replaced the endogenous T cell receptor (TCR) locus with a new TCR that redirected T cells to a cancer antigen. The resulting TCR-engineered T cells specifically recognized tumour antigens and mounted productive anti-tumour cell responses in vitro and in vivo. Together, these studies provide preclinical evidence that non-viral genome targeting can enable rapid and flexible experimental manipulation and therapeutic engineering of primary human immune cells.

A major barrier to effective non-viral T cell genome targeting of large DNA sequences has been the toxicity of the DNA⁷. Although the introduction of short single-stranded oligodeoxynucleotide (ssODN) homology-directed repair (HDR) templates does not cause notable T cell toxicity, it has been shown that larger linear double-stranded DNA (dsDNA) templates are toxic at high concentrations^{8,9}. Contrary to expectations, we found that co-electroporation of human primary T cells with CRISPR–Cas9 ribonucleoprotein (RNP)^{10,11} complexes and long (>1 kb) linear dsDNA templates reduced the toxicity associated with the dsDNA template (Extended Data Fig. 1a–e). Cas9 RNPs were co-electroporated with a dsDNA HDR template designed to introduce

an N-terminal green fluorescent protein (GFP) fusion in the house-keeping gene *RAB11A* (Fig. 1a). Both cell viability and the efficiency of this approach were optimized by systematic exploration (Fig. 1b and Extended Data Fig. 1f–h), resulting in GFP expression in up to 50% of primary human CD4⁺ and CD8⁺ T cells. The method was reproducibly efficient with high cell viability (Fig. 1c–e). The system is also compatible with current manufacturing protocols for cell therapies. The method can be used with fresh or cryopreserved cells, bulk T cells or sub-populations sorted by fluorescence activated cell sorting (FACS), and cells from whole blood or leukapheresis (Extended Data Fig. 2a–d).

We next confirmed that the system could be applied broadly by targeting sequences in different locations throughout the genome. We efficiently engineered primary T cells by generating GFP fusions with different genes (Fig. 2a and Extended Data Fig. 2e-g). Live-cell imaging with confocal microscopy confirmed the specificity of gene targeting, revealing the distinct sub-cellular locations of each of the resulting GFP-fusion proteins¹² (Fig. 2b). Appropriate chromatin binding of a transcription factor GFP-fusion protein was confirmed by performing genome-wide CUT&RUN (cleavage under targets and release using nuclease)¹³ analysis with an anti-GFP antibody (Fig. 2c and Extended Data Fig. 2h). Finally, we showed that gene targeting preserved the regulation of the modified endogenous gene. Consistent with correct cell-type specific expression, a CD4-GFP fusion was selectively expressed in the CD4⁺ population of T cells (Fig. 2d). Using HDR templates encoding several fluorescent proteins, we demonstrated that we could generate cells with bi-allelic gene targeting (Fig. 2e and Extended Data Fig. 3a-d) or multiplex modification of two (Fig. 2f and Extended Data Fig. 3e-h) or even three (Fig. 2g and Extended Data Fig. 3i) different genes^{14,15}. These results show that several endogenous genes can be directly engineered without virus in T cells, and that gene and protein regulation are preserved.

For the rapeutic use of genetically modified T cells, integrated sequences should be introduced specifically without unintended disruption of other critical genome sites¹⁶. We performed targeted locus

¹Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA, USA. ²Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA. ³Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA. ⁶Diabetes Center, University of California, San Francisco, San Francisco, CA, USA. ⁶Department of Medicine, University of California at Los Angeles, Los Angeles, CA, USA. ⁷UCSF Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA. ⁸Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. ⁸Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. ¹⁰HIV Dynamics and Replication Program, Vector Design and Replication Section, National Cancer Institute, Frederick, MD, USA. ¹¹Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. ¹²Division of Stem Cell Transplantation and Regenerative Medicine, Department of Pediatrics, Stanford University, Stanford, CA, USA. ¹³Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, USA. ¹⁴Section of Adult and Pediatric Endocrinology, Diabetes, and Metabolism, Departments of Medicine and Pediatrics, The University of Chicago, Chicago, IL, USA. ¹⁵Department of Human Genetics, The University of Chicago, Chicago, IL, USA. ¹⁶Takara Bio USA, Inc, Mountain View, CA, USA. ¹⁷Chan Zuckerberg Biohub, San Francisco, CA, USA. ¹⁸Mouse Genome Engineering Core Facility, Vice Chancellor for Research Office, University of Nebraska Medical Center, Omaha, NE, USA. ¹⁹Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA. ²⁰Department of Pediatrics, Pathology, Yale School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ²³Department of Immunobiology and Intern

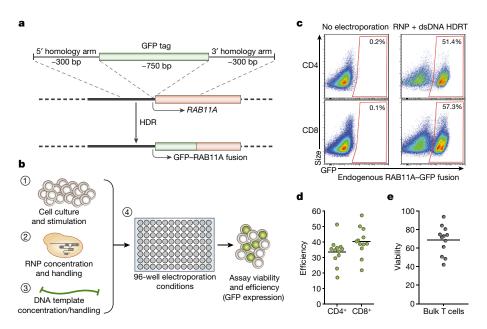


Fig. 1 | Efficient non-viral genome targeting in primary human T cells. a, HDR-mediated integration of a GFP fusion tag to the housekeeping gene RAB11A. b, Development and optimization of non-viral genome targeting for both cell viability and HDR efficiency. c, Insertion of a GFP fusion into the endogenous RAB11A gene using non-viral targeting in primary human gated CD4+ and CD8⁺ T cells. HDRT, HDR template. **d**, Average efficiency with the RAB11A-GFP HDR template was 33.7% and 40.3% in CD4⁺ and CD8⁺ cells, respectively. **e**, Viability (number of live cells relative to non-electroporated control) after non-viral genome targeting averaged 68.6%. Efficiency and viability were measured 4 days after electroporation. Mean values of n = 12 independent healthy donors are shown (horizontal bars, d, e). See also Extended Data Fig. 1.

amplification (TLA) sequencing¹⁷ and found no evidence of off-target integrations above the limit of detection of the assay (approximately 1% of alleles) (Extended Data Fig. 4a, b). We further assessed potential

off-target integrations at the single-cell level by quantifying GFP⁺ cells generated using a Cas9 RNP that cuts outside the homology site. Similar to what has been described with viral HDR templates^{5,18}, we

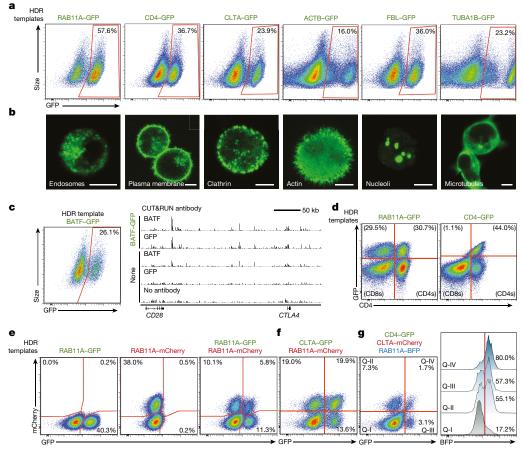


Fig. 2 | Individual and multiplexed modification of endogenous T cell genes. a, Non-viral genome targeting with GFP-fusion constructs into several endogenous genes. b, Confocal microscopy of live human T cells electroporated with the indicated HDR templates confirmed fusion-protein localization. Scale bars, 5 $\mu m.$ c, GFP fused to the endogenous transcription factor BATF enabled genome-wide binding analysis (CUT&RUN) using anti-GFP or anti-BATF antibodies. d, RAB11A-fusions produced GFP-positive CD4+ and CD8+ cells, whereas the CD4-

fusions were selectively expressed in CD4⁺ cells. **e**, Bi-allelic non-viral genome targeting of two distinct fluorescent proteins into the same locus. **f**, Multiplexed non-viral genome targeting of HDR templates into two separate genomic loci. **g**, Simultaneous targeting of three distinct genomic loci. Cells positive for one (Q-II, Q-III) or two (Q-IV) integrations were highly enriched for a third HDR integration. BFP, blue fluorescent protein. One representative donor displayed from n=6 (**a**), n=4 (**b**, **d**–**g**), or n=2 (**c**) independent healthy donors. See also Extended Data Figs. 2, 3.

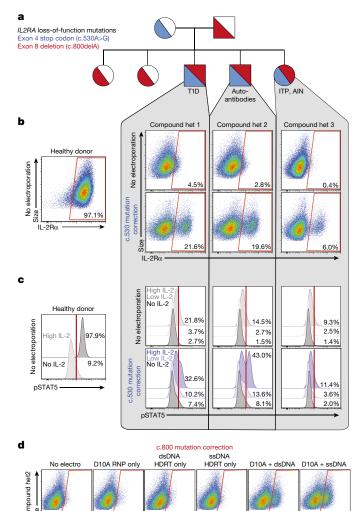


Fig. 3 | Monogenic autoimmune mutations corrected by non-viral genome targeting. a, Pedigree of family with monogenic immune disease caused by compound heterozygous (het) mutations in IL2RA (Supplementary Table 4). **b**, Correction of c.530A>G *IL2RA* mutation by non-viral genome targeting in three compound heterozygous siblings rescued IL-2R α cell surface expression on CD3⁺ T cells 2 days after electroporation. AIN, autoimmune neutropenia; ITP, idiopathic thrombocytopenic purpura. c, Seven days after non-viral genome targeting, targeted unselected CD3+T cells showed increased STAT5 phosphorylation (pSTAT5) levels after IL-2 stimulation compared to nontargeted controls. d, Non-viral genome targeting corrected the c.800delA mutation using D10A nickase and a long ssDNA HDR template. IL-2R α surface expression was measured after 9 days of ex vivo expansion following electroporation (2 days after re-stimulation). n = 3 (**b**, **c**) or n = 1 (**d**) compound heterozygous patients per correction. See also Extended Data Figs. 6–8.

found evidence to suggest that double-stranded templates could integrate independent of target homology 19,20 , albeit at low rates (Extended Data Fig. 4c–i). These rare events could be reduced almost completely by using single-stranded DNA (ssDNA) templates 21,22 (Extended Data Fig. 5a–d). As an additional safeguard that could be important for some applications, we demonstrated that efficient non-viral T cell genome targeting also could be achieved using either a single-stranded or a double-stranded template with a Cas9 'nickase' engineered to reduce potential off-target double-stranded cuts 23,24 (Extended Data Fig. 5e–h).

Having optimized this non-viral genome engineering approach in primary human T cells, we demonstrated its use in two different clinically relevant settings in which the targeted replacement of a gene would provide proof-of-principle that the method can be used to create

therapeutically relevant gene modifications. Specifically, we tested the ability to rapidly and efficiently correct an inherited genetic alteration in T cells, and we also tested the targeted insertion of the two chains of a TCR to redirect the specificity of T cells to recognize cancer cells.

We identified a family with monogenic primary immune deficiency with autoimmune disease caused by recessive loss-of-function mutations in the gene encoding the IL-2 α receptor (IL2RA)²⁵ (Supplementary Table 4), which is essential for healthy regulatory T (T_{reg}) cells²⁶ (Extended Data Fig. 6a-h). Whole-exome sequencing revealed that the IL2RA-deficient children contained compound heterozygous mutations in IL2RA (Fig. 3a and Extended Data Fig. 6i). One mutation, c.530A>G, creates a premature stop codon. With non-viral genome targeting, we were able to correct the mutation and observed IL-2R α expression on the surface of corrected T cells from the patient (Fig. 3b). Long dsDNA templates led to efficient correction of the mutations. Because only two base pair changes were necessary (one to correct the mutation and one to silently remove the PAM sequence of the guide RNA (gRNA)), a short ssDNA (approximately 120 base pairs (bp)) could also be used to make the correction. These ssDNAs were able to correct the mutation at high frequencies, although here the efficiency of correction was lower than with the longer dsDNA template (Extended Data Figs. 7a, 8a). Correction was successful in T cells from all three siblings, but lower rates of IL-2R α expression were seen in compound heterozygote 3, which could be due to altered cell-state associated with the patient's disease or the fact that she was the only sibling treated with immunosuppressive therapy (Supplementary Table 4 and Extended Data Fig. 8f). The second mutation identified, c.800delA, causes a frameshift in the reading frame of the final *IL2RA* exon. This frameshift mutation could be corrected both by HDR as well as by RNP cutting alone, presumably owing to some of the small indels restoring the reading frame (Extended Data Fig. 8). Together, these data show that distinct mutations can be corrected in patient T cells using HDR template-dependent and non-HDR template-dependent mechanisms.

Mutation correction improved cell signalling function. After correction of the c.530A>G IL2RA mutation, treatment with IL-2 led to increased phosphorylation of STAT5, a hallmark of productive signalling (Fig. 3c and Extended Data Figs. 7c, 8c). In addition, after correction, we found that the modified T cells expressed both IL-2R α and FOXP3, a crucial transcriptional factor in T_{reg} cells (Extended Data Figs. 7d, 8d). We were also able to correct the IL2RA mutation in a sorted population of CD3⁺CD4⁺CD127^{lo}TIGIT⁺CD45RO⁺ T_{reg}-like cells from a patient (Extended Data Fig. 7e, f), a strategy that could potentially be used in a gene-modified cell therapy for the children in this family. Cell-type specific and stimulus responsive expression of IL2RA is under tight control by multiple endogenous cis-regulatory elements that constitute a super-enhancer^{27,28}. Therefore, effective therapeutic correction of the IL2RA defect is likely to depend on repairing the gene in its endogenous genomic locus; off-target effects should be avoided. We therefore demonstrated that the c.800delA mutation could also be repaired using Cas9 nickase combined with a single-stranded HDR template (Fig. 3d).

Non-viral genome targeting not only allows the correction of point mutations, but also enables integration of much larger DNA sequences. We were able to use a large DNA construct to rapidly reprogram the antigen specificity of human T cells, which is critical for many cellular immunotherapy applications. Recent work demonstrates that chimaeric antigen receptors (CARs) have enhanced efficacy when they are genetically encoded in the endogenous TCR locus using CRISPR–Cas9 gene cutting and an adeno-associated virus vector as a repair template 5. Targeting of specific TCR sequences to this locus is a more challenging problem because T cells must express paired TCR alpha (TCR- α) and beta chains (TCR- β) to make a functional receptor.

We developed a strategy to replace the endogenous TCR using non-viral genome targeting to integrate an approximately 1.5-kb DNA cassette into the first exon of the TCR- α constant region (TRAC) (Fig. 4a). This cassette encoded the full-length sequence of a TCR- β separated by a self-excising 2A peptide from the variable region of a

RESEARCH LETTER

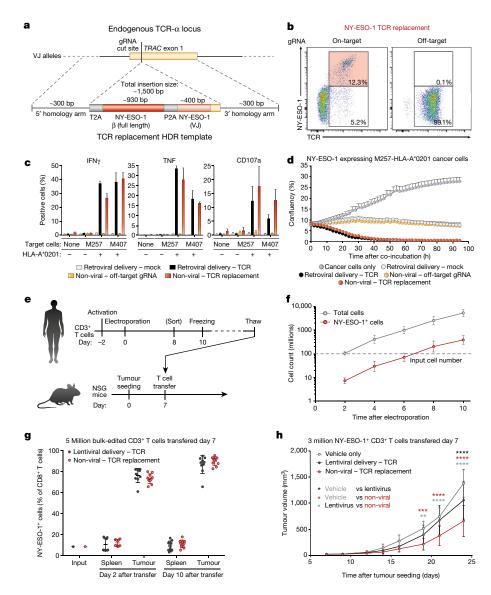


Fig. 4 | Replacement of the endogenous TCR by non-viral genome targeting. a, Schematic of HDR template used to replace the endogenous TCR. b, Non-viral genome targeting successfully replaced the endogenous TCR with the NY-ESO-1 antigen-specific 1G4 TCR. c, Antigen-specific cytokine production and degranulation in CD8⁺ T cells with the replaced TCR. d, Antigen-specific target cell killing by CD8⁺ T cells with the replaced TCR. e, Melanoma tumour mouse xenograft model. NSG mice, non-obese diabetic (NOD)/severe combined immunodeficiency (SCID)/Il2rg^{-/-} mice. f, Scalability of non-viral replacement of the endogenous TCR for adoptive cell therapy. g, Preferential in vivo localization of NY-ESO-1 TCR+ T cells to the tumour. h, Tumour growth after adoptive transfer of NY-ESO-1 TCR⁺ non-virally or lentivirally modified or vehicle alone (saline). One representative donor from n = 6 (**b**) or n = 2(c, d) independent healthy donors, with mean and s.d. of technical triplicates (c, d). n = 6 (f)or n = 2 (**g**, **h**) independent healthy donors in 5 (g) or 7 mice (h) with mean and s.d. (f-h). **P < 0.01, ***P < 0.001, ****P < 0.0001 (two-way analysis of variance (ANOVA) with Holm-Sidak's multiple comparisons test). See also Extended Data Figs. 9, 10.

new TCR- α , which encodes the full TCR- α sequence when appropriately integrated at the endogenous TRAC exon (Extended Data Fig. 9a–d). To test this strategy, we introduced a TCR- β and TCR- α pair (1G4) that recognizes the NY-ESO-1 tumour antigen²⁹ into the TRAC locus of polyclonal T cells isolated from healthy human donors. Antibody staining for total TCR- α/β expression and NY-ESO-1-MHC dextramer staining for the NY-ESO-1 TCR expression revealed that non-viral genome targeting enabled reproducible replacement of the endogenous TCR in both CD8⁺ and CD4⁺ primary human T cells (Fig. 4b and Extended Data Fig. 9k). NY-ESO-1 TCR cells could also be generated with a similar targeting strategy at the TCR-β constant region (TRBC1/2) or with multiplexed simultaneous replacement of both endogenous TCR- α and TCR- β (Extended Data Fig. 9e–i). Most of the T cells that did not express NY-ESO-1 TCR were TCR knockouts (Fig. 4b), presumably due to non-homologous end joining (NHEJ) events induced by the Cas9-mediated double-stranded DNA breaks in TRAC exon 1. Up to around 70% of resulting TCR-positive cells recognized the NY-ESO-1 dextramer.

Next, we assessed the tumour antigen-specific function of targeted human T cells. When the targeted T cells were co-cultured with two different NY-ESO-1 $^+$ melanoma cell lines, M257 and M407, the modified T cells robustly and specifically produced IFN γ and TNF and induced T cell degranulation (measured by CD107a surface expression) (Fig. 4c). Cytokine production and degranulation only occurred when the NY-ESO-1 TCR T cells were exposed to cell lines expressing the

appropriate human leukocyte antigen (HLA)-A*0201 class I major histocompatibility complex (MHC) allele required to present the cognate NY-ESO-1 peptide to the TCR. Both the CD8⁺ and CD4⁺ T cell response was consistent across healthy donors, and was comparable to the response of T cells from the same healthy donor in which the NY-ESO-1 TCR was transduced by gamma retrovirus and heterologously expressed using a viral promoter (Fig. 4c and Extended Data Fig. 9j). NY-ESO-1 TCR knock-in T cells rapidly killed target M257-HLA-A*0201 cancer cells in vitro at rates similar to the positive control, retrovirally transduced T cells (Fig. 4d). Killing was selective for target cells expressing NY-ESO-1 antigen and the HLA-A*0201 allele, consistent across donors, and depended on the T cells being modified using both the correct gRNA and HDR template (Extended Data Fig. 9n–q).

Finally, we confirmed that non-viral genome targeting could be used to generate NY-ESO-1 TCR cells at scale and that these cells have in vivo anti-tumour function (Fig. 4e and Extended Data Fig. 10a). Given that knock-in efficiency was lower with non-viral targeting than with comparable sized adeno-associated virus templates⁵, we first wanted to ensure that we could generate sufficient numbers of NY-ESO-1-positive cells for adoptive cell therapies. We electroporated 100 million T cells from six healthy donors, which after ten days of expansion yielded an average of 385 million NY-ESO-1 TCR T cells per donor (Fig. 4f and Extended Data Fig. 9i-m). NY-ESO-1 TCR knock-in T cells preferentially localized to, persisted at, and proliferated in the tumour rather than the spleen, similar to positive control lentivirally-transduced

T cells (Fig. 4g and Extended Data Fig. 10b-f). Adoptive transfer of sorted NY-ESO-1 TCR T cells also reduced the tumour burden in treated animals (Fig. 4h).

Our therapeutic gene editing in human T cells is a process that takes only a short time from target selection to production of the genetically modified T cell product. In approximately one week, novel gRNAs and DNA repair templates can be designed, synthesized, and the DNA integrated into primary human T cells that remain viable, expandable, and functional. The whole process and all required materials can be easily adapted to good manufacturing practices for clinical use. Avoiding the use of viral vectors will accelerate research and clinical applications, reduce the cost of genome targeting, and potentially improve safety.

Looking forward, the technology could be used to 'rewire' complex molecular circuits in human T cells. Multiplexed integration of large functional sequences at endogenous loci should allow combinations of coding and non-coding elements to be corrected, inserted, modified and rearranged. Much work remains to be done to improve our understanding of endogenous T cell circuitry if we are going to create synthetic circuits. Rapid and efficient non-viral tagging of endogenous genes in primary human cells will facilitate live-cell imaging and proteomic studies to decode T cell programs. Non-viral genome targeting provides an approach to re-write these programs in cells for the next generation of immunotherapies.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0326-5.

Received: 19 November 2017; Accepted: 4 June 2018; Published online 11 July 2018.

- Sadelain, M., Rivière, I. & Riddell, S. Therapeutic T cell engineering. Nature 545, 423–431 (2017).
- Lim, W. A. & Juné, C. H. The principles of engineering immune cells to treat cancer. Cell. 168, 724–740 (2017).
- Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. Science 348, 62–68 (2015).
- Verhoeyen, E., Costa, C. & Cosset, F.-L. Lentiviral vector gene transfer into human T cells. Methods Mol. Biol. 506, 97–114 (2009).
- Eyquem, J. et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* 543, 113–117 (2017).
- Hale, M. et al. Homology-directed recombination for enhanced engineering of chimeric antigen receptor T cells. Mol. Ther. Methods Clin. Dev. 4, 192–203 (2017).
- Cornu, T. I., Mussolino, C. & Cathomen, T. Refining strategies to translate genome editing to the clinic. *Nat. Med.* 23, 415–423 (2017).
 Zhao, Y. et al. High-efficiency transfection of primary human and mouse T
- 8. Zhao, Y. et al. High-efficiency transfection of primary human and mouse lymphocytes using RNA electroporation. *Mol. Ther.* **13**, 151–159 (2006).
- Hornung, V. & Latz, E. Intracellular DNA recognition. Nat. Rev. Immunol. 10, 123–130 (2010).
- Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24, 1012–1019 (2014).
- Schumann, K. et al. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc. Natl Acad. Sci. USA 112, 10437–10442 (2015).
- Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl Acad. Sci. USA* 113, E3501–E3508 (2016).
- Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for highresolution mapping of DNA binding sites. eLife 6, e21856 (2017).
- Bak, R. O. et al. Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. eLife 6, e27873 (2017).
- 15. Agudelo, D. et al. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods* **14**, 615–620 (2017).
- Lux, C. T. & Scharenberg, A. M. Therapeutic gene editing safety and specificity. Hematol. Oncol. Clin. North Am. 31, 787–795 (2017).
- Cain-Hom, C. et al. Efficient mapping of transgene integration sites and local structural changes in Cre transgenic mice using targeted locus amplification. Nucleic Acids Res. 45, e62 (2017).
- Dever, D. P. et al. CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. Nature 539, 384–389 (2016).

- Murnane, J. P., Yezzi, M. J. & Young, B. R. Recombination events during integration of transfected DNA into normal human cells. *Nucleic Acids Res.* 18, 2733–2738 (1990).
- Suzuki, K. et al. *In vivo* genome editing via CRISPR/Cas9 mediated homologyindependent targeted integration. *Nature* 540, 144–149 (2016).
- Quadros, R. M. et al. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol. 18, 92 (2017).
- 22. Li, H. et al. Design and specificity of long ssDNA donors for CRISPR-based knock-in. Preprint at https://doi.org/10.1101/178905 (2017).
- Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963 (2013).
- Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013).
- Sharfe, N., Dadi, H. K., Shahar, M. & Roifman, C. M. Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. *Proc. Natl Acad. Sci. USA* 94, 3168–3171 (1997).
- Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J. Immunol. 155, 1151–1164 (1995).
- Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343 (2015).
- Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* 549, 111–115 (2017).
- Robbins, P. F. et al. Single and dual amino acid substitutions in TCR CDRs can enhance antigen-specific T cell functions. J. Immunol. 180, 6116–6131 (2008).

Acknowledgements We thank members of the Marson laboratory, C. Jeans, K. Marchuk, J. Bluestone, Q. Tang, R. Wagner, the UCSF Biological Imaging Development Center, the UCSF Parnassus Center for Advanced Technology, the UCSF Parnassus Flow Cytometry Core (NIH P30 DK063720 and 1S100D021822-01), Lonza, J. Corn and S. Pyle for suggestions and assistance. This research was supported by NIH grants DP3DK111914-01 (A.M.), P50GM082250 (A.M.), R35 CA197633 (A.R.), K23 DK094866 (S.W.G.), T32GM007618 (T.L.R., J.H.), T32 DK007418 (T.L.R.), and P30 DK020595 (S.W.G.), the NIH NCI Intramural Program (A.L.F., S.H.H.), grants from the Keck Foundation (A.M.), National Multiple Sclerosis Society (A.M.; CA 1074-A-21), gifts from J. Aronov, G. Hoskin, the Jeffrey Modell Foundation (A.M.), and awards from the Burroughs Wellcome Fund (A.M.) and the Ressler Family Fund (C.P.S., J.S., A.R.). A.M. is a Chan Zuckerberg Biohub investigator. A.R. is a Parker Institute for Cancer Immunotherapy member.

Reviewer information *Nature* thanks M. Maus, J. Wherry and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.L.R. and A.M. designed the study and wrote the manuscript. T.L.R. designed and performed all electroporation experiments. T.L.R., R.Y., E.S., J.L., J.H., V.T., D.M.N. and K.S. contributed to functional assays of edited T cells. R.Y. performed and analysed CUT&RUN experiments. H.L., J.W. and M.D.L. developed the IVT–RT ssDNA production method. H.M., M.M., Y.M., B.S. and M.H. developed the exonuclease-based ssDNA production method. R.Q. and C.G. discussed the use of ssDNA. A.M.F. and S.H.H. advised on methods of DNA introduction into T cells. T.L.R., E.S., M.C. and A.P.M. performed amplicon sequencing. J.C., J.N.S., A.L.P., L.P., D.C., G.A.A., D.D.G., G.M.K., S.W.G., R.B., E.M., M.G.R., N.R. and K.C.H. contributed to the clinical workup of *IL2RA*-deficient family and functional assays on unedited patient T cells. J.H.E. and M.R.L. performed TSDR analysis. T.L.R., C.P.S., E.S., A.R. and A.M. designed the endogenous TCR knock-in strategy. T.L.R., C.P.S., J.C., J.S., P.K., A.A. and A.R. performed or supervised in vitro assays of T cells with endogenous TCR knock-ins. T.L.R. designed and performed all mouse experiments.

Competing interests A.M. is a co-founder of Spotlight Therapeutics. A.M. serves as an advisor to Juno Therapeutics and is a member of the scientific advisory board of PACT Pharma. The Marson laboratory has received sponsored research support (Juno Therapeutics, Epinomics, Sanofi) and a gift from Gilead. A.R. is co-founder and a member of the scientific advisory board of PACT Pharma. T.L.R., C.P.S., E.S., A.R. and A.M. are inventors on new patent applications related to this manuscript (US patent application no. 62/520,117, T.L.R. and A.M.; US patent application no. 62/578,153, T.L.R., C.P.S., E.S., A.R. and A.M.).

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0326-5

 $\label{eq:supplementary information} Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0326-5.$

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.M. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. For all in vivo experiments, experimental conditions were allocated randomly at the time of adoptive transfer, and experimental conditions were mixed among littermates. For in vivo tumour sizing experiments, the investigator was blinded to experimental condition. No power analysis was used to determine sample sizes. Antibodies. All antibodies used in the study for fluorescence activated cell sorting, flow cytometry and cellular stimulations are listed in Supplementary Table 2. **Guide RNAs.** All gRNAs used in the study are listed in Supplementary Table 3. Isolation of human primary T cells for gene targeting. Primary human T cells were isolated from healthy human donors either from fresh whole blood, residuals from leukoreduction chambers after Trima Apheresis (Blood Centers of the Pacific), or leukapheresis products (StemCell). Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples by Ficoll centrifugation using SepMate tubes (STEMCELL, per manufacturer's instructions). T cells were isolated from PBMCs from all cell sources by magnetic negative selection using an EasySep Human T Cell Isolation Kit (STEMCELL, per manufacturer's instructions). Unless otherwise noted, isolated T cells were stimulated as described below and used directly (fresh). When frozen cells were used, previously isolated T cells that had been frozen in Bambanker freezing medium (Bulldog Bio) per manufacturer's instructions were thawed, cultured in media without stimulation for 1 day, and then stimulated and handled as described for freshly isolated samples. Fresh blood was taken from healthy human donors under a protocol approved by the UCSF Committee on Human Research (CHR #13-11950). Patient samples used for gene editing were obtained under a protocol approved by the Yale Human Investigation Committee (HIC). Additional leukapheresis products from healthy donors were collected either under UCLA Institutional Review Board (IRB) approval #10-001598 or purchased from AllCells, LLC. All patients and healthy donors provided informed consent.

Primary human T cell culture. Unless otherwise noted, bulk T cells were cultured in XVivo15 medium (STEMCELL) with 5% fetal bovine serum (FBS), 50 μM 2-mercaptoethanol, and 10 μM *N*-acetyl L-cystine. Immediately after isolation, T cells were stimulated for 2 days with anti-human CD3/CD28 magnetic dynabeads (ThermoFisher) at a beads to cells concentration of 1:1, along with a cytokine cocktail of IL-2 at 200 U ml $^{-1}$ (UCSF Pharmacy), IL-7 at 5 ng ml $^{-1}$ (ThermoFisher), and IL-15 at 5 ng ml $^{-1}$ (Life Tech). After electroporation, T cells were cultured in media with IL-2 at 500 U ml $^{-1}$. Throughout the culture period T cells were maintained at an approximate density of 1 million cells per ml of media. Every 2–3 days after electroporation, additional media was added, along with additional fresh IL-2 to bring the final concentration to 500 U ml $^{-1}$, and cells were transferred to larger culture vessels as necessary to maintain a density of 1 million cells per ml.

RNP production. RNPs were produced by complexing a two-component gRNA to Cas9, as previously described 11 . In brief, crRNAs and tracrRNAs were chemically synthesized (Dharmacon, IDT), and recombinant Cas9-NLS, D10A-NLS, or dCas9-NLS were recombinantly produced and purified (QB3 Macrolab). Lyophilized RNA was resuspended in 10 mM Tris-HCL (7.4 pH) with 150 mM KCl at a concentration of 160 μ M, and stored in aliquots at $-80\,^{\circ}\mathrm{C}$. crRNA and tracrRNA aliquots were thawed, mixed 1:1 by volume, and annealed by incubation at $37\,^{\circ}\mathrm{C}$ for 30 min to form an 80 μ M gRNA solution. Recombinant Cas9 or the D10A Cas9 variant were stored at 40 μ M in 20 mM HEPES-KOH, pH 7.5, 150 mM KCl, 10% glycerol, 1 mM DTT, were then mixed 1:1 by volume with the 80 μ M gRNA (2:1 gRNA to Cas9 molar ratio) at $37\,^{\circ}\mathrm{C}$ for 15 min to form an RNP at 20 μ M. RNPs were electroporated immediately after complexing.

Double-stranded DNA HDRT production. Novel HDR sequences were constructed using Gibson Assemblies to insert the HDR template sequence, consisting of the homology arms (commonly synthesized as gBlocks from IDT) and the desired insert (such as GFP) into a cloning vector for sequence confirmation and future propagation. These plasmids were used as templates for high-output PCR amplification (Kapa Hotstart polymerase). PCR amplicons (the dsDNA HDRT) were SPRI purified (1.0 \times) and eluted into a final volume of 3 μ l H₂O per 100 μ l of PCR reaction input. Concentrations of HDRTs were determined by nanodrop using a 1:20 dilution. The size of the amplified HDRT was confirmed by gel electrophoresis in a 1.0% agarose gel. All homology directed repair template sequences used in the study, both dsDNA and ssDNA, are listed in Supplementary Table 3. Single-stranded DNA HDRT production by exonuclease digestion. To produce long ssDNA as HDR templates, the DNA of interest was amplified via PCR using one regular, non-modified PCR primer and a second phosphorylated PCR primer. The DNA strand that will be amplified using the phosphorylated primer will be the strand that will be degraded using this method. This makes it possible to prepare either a single-stranded sense or single-stranded antisense DNA using the respective phosphorylated PCR primer. To produce the ssDNA strand of interest, the phosphorylated strand of the PCR product was degraded by treatment with two enzymes, Strandase Mix A and Strandase Mix B, for 5 min (per 1 kb) at 37 °C, respectively. Enzymes were deactivated by a 5 min incubation at 80 °C. The resulting ssDNA HDR templates were SPRI purified (1.0X) and eluted in H_2O . A more detailed protocol for the Guide-it Long ssDNA Production System (Takara Bio, 632644) can be found at the manufacturer's website.

Single-stranded DNA HDRT production by reverse synthesis. ssDNA HDR templates were synthesized by reverse transcription of an RNA intermediate followed by hydrolysis of the RNA strand in the resulting RNA:DNA hybrid product, as described²². In brief, the desired HDR donor was first cloned downstream of a T7 promoter and the T7-HDR donor sequence amplified by PCR. RNA was synthesized by in vitro transcription using HiScribe T7 RNA polymerase (New England Biolabs) and reverse-transcribed using TGIRT-III (InGex). After reverse transcription, NaOH and EDTA were added to 0.2 M and 0.1 M, respectively, and RNA hydrolysis was carried out at 95 °C for 10 min. The reaction was quenched with HCl, the final ssDNA product purified using Ampure XP magnetic beads (Beckman Coulter) and eluted in sterile RNase-free H2O. ssDNA quality was analysed by capillary electrophoresis (Bioanalyzer, Agilent).

Primary T cell electroporation. RNPs and HDR templates were electroporated 2 days after initial T cell stimulation. T cells were collected from their culture vessels and magnetic anti-CD3/anti-CD28 dynabeads were removed by placing cells on an EasySep cell separation magnet for 2 min. Immediately before electroporation, de-beaded cells were centrifuged for 10 min at 90g, aspirated, and resuspended in the Lonza electroporation buffer P3 using 20 μ l buffer per 1 million cells. For optimal editing, 1 million T cells were electroporated per well using a Lonza 4D 96-well electroporation system with pulse code EH115. Alternate cell concentrations from 200,000 up to 2 million cells per well resulted in lower transformation efficiencies. Alternate electroporation buffers were used as indicated, but had different optimal pulse settings (EO155 for OMEM buffer). Unless otherwise indicated, 2.5 μ l RNPs (50 pmol total) were electroporated, along with 2 μ l HDR Template at 2 μ g μ l $^{-1}$ (4 μ g HDR template total).

The order of cell, RNP and HDRT addition appeared to matter (Extended Data Fig. 1). For 96-well experiments, HDRTs were first aliquoted into wells of a 96-well polypropylene V-bottom plate. RNPs were then added to the HDRTs and allowed to incubate together at room temperature for at least 30 s. Finally, cells resuspended in electroporation buffer were added, briefly mixed by pipetting with the HDRT and RNP, and 24 μl of total volume (cells plus RNP and HDRT) was transferred into a 96-well electroporation cuvette plate. Immediately after electroporation, 80 µl of pre-warmed media (without cytokines) was added to each well, and cells were allowed to rest for 15 min at 37 °C in a cell culture incubator while remaining in the electroporation cuvettes. After 15 min, cells were moved to final culture vessels. Flow cytometry and cell sorting. Flow cytometric analysis was performed on an Attune NxT Acoustic Focusing Cytometer (ThermoFisher) or an LSRII flow cytometer (BD). FACS was performed on the FACSAria platform (BD). Surface staining for flow cytometry and cell sorting was performed by pelleting cells and resuspending in 25 µl of FACS buffer (2% FBS in PBS) with antibodies at the indicated concentrations (Supplementary Table 2) for 20 min at 4 °C in the dark. Cells were washed once in FACS buffer before resuspension.

Confocal microscopy. Samples were prepared by drop casting 10 µl of a solution of suspended live T cells onto a 3×1 inch $(7.6\times2.5$ cm) microscope slide onto which a 25 mm² coverslip was placed. Imaging was performed on an upright configuration Nikon A1r laser scanning confocal microscope. Excitation was achieved through a 488 nm OBIS laser (Coherent). A long working distance (LWD) $60\times$ Plan Apo 1.20 numerical aperture (NA) water immersion objective was used with additional digital zoom achieved through the NIS-Elements software. Images were acquired under 'Galvano' mirror settings with $2\times$ line averaging enabled and exported as TIFF to be analysed in FIJI (ImageJ, NIH).

CUT&RUN. CUT&RUN was performed using epitope-tagged primary human T cells 11 days after electroporation and 4 days after re-stimulation with anti-CD3/ anti-CD28 dynabeads (untagged cells were not electroporated). Approximately 20% and 10% of electroporated cells showed GFP-BATF expression as determined by flow cytometry in donor 1 and donor 2 samples, respectively. CUT&RUN was performed as described¹³, using anti-GFP (ab290), anti-BATF (sc-100974), and rabbit anti-mouse (ab46540) antibodies. In brief, 6 million cells (30 million cells for anti-GFP CUT&RUN in GFP-BATF-containing cells) were collected and washed. Nuclei were isolated and incubated rotating with primary antibody (GFP or BATF) for 2 h at 4 °C. BATF CUT&RUN samples were incubated for an additional hour with rabbit anti-mouse antibody. Next, nuclei were incubated with proteinAmicrococcal nuclease (provided by the Henikoff laboratory) for 1 h at 4 °C. Nuclei were equilibrated to 0 °C and MNase digestion was allowed to proceed for 30 min. Solubilized chromatin CUT&RUN fragments were isolated and purified. Paired-end sequencing libraries were prepared and analysed on Illumina Nextseq machines and sequencing data were processed as described¹³. For peak calling and heatmap generation, reads mapping to centromeres were filtered out.

TLA sequencing and analysis. TLA sequencing was performed by Cergentis as previously described¹⁷. Similarly, data analysis of integration sites and transgene fusions was performed by Cergentis as previously described¹⁷. TLA sequencing was performed in two healthy donors, each edited at the *RAB11A* locus with either

a dsDNA or ssDNA HDR template to integrate a GFP fusion (Fig. 1b). Sequencing reads showing evidence of primer dimers or primer bias (that is, greater than 99% of observed reads came from single primer set) were removed.

In vitro T_{reg} cell suppression assay. $CD4^+$ T cells were enriched using the EasySep Human $CD4^+$ T cell enrichment kit (STEMCELL Technologies). $CD3^+CD4^+CD127^{lo}CD45RO^+TIGIT^+$ enriched T_{reg} -like cells from IL2RA-deficient subjects and healthy donors as well as $CD3^+CD4^+IL-2R\alpha^{hi}CD127^{lo}$ T_{reg} cells from $IL2RA^{+/-}$ heterozygous individuals were sorted by flow cytometry. $CD3^+CD4^+IL-2R\alpha^-CD127^+$ responder T (T_{resp}) cells were labelled with CellTrace CFSE (Invitrogen) at 5 μ M. T_{reg} cells and healthy donor T_{resp} cells were cocultured at a 1:1 ratio in the presence of beads loaded with anti-CD2, anti-CD3 and anti-CD28 (Treg Suppression Inspector; Miltenyi Biotec) at a 1 bead: 1 cell ratio. On days 3.5–4.5, co-cultures were analysed by FACS for CFSE dilution. The percentage inhibition is calculated using the following formula: 1- (% proliferation with T_{reg} cells/% proliferation of stimulated T_{resp} cells without T_{reg} cells).

Sorting and TSDR analysis of corrected T_{reg} cells. Ex vivo expanded T_{reg} and T effector cells from a healthy control and a patient with IL2RA compound heterozygous mutations (D6) were thawed and stained. Live cells were sorted based on expression of CD25 and CD62L markers directly into ZymoResearch M-digestion Buffer (2x) (D5021-9) supplemented with proteinase K. The lysate was incubated at 65 °C for greater than 2 h and then frozen. Bisulfite conversion and pyrosequencing of the samples was performed by EpigenDx (assay ID ADS783-FS2) to interrogate the methylation status of 9 CpG sites intron 1 of the FOXP3 gene, spanning -2330 to -2263 from ATG.

Generation of retrovirally and lentivirally transduced control T cells. For retroviral infections, clinical grade MSGV-1-1G4 (NY-ESO-1 TCR transgene) retroviral vector (IUVPC) was used. For lentiviral production, HEK293 cells were plated at 18 million cells in 15 cm dishes the night before transfection. Cells were transfected using the Lipofectamine 3000 reagent following the manufacturer's protocol (L3000001). Transfection media was changed the following day to fresh HEK293 media (DMEM with 5% FBS and 1% penicillin/streptromycin) with viral boost reagent per the manufacturer's protocol at $500\times$ (Alstem viral boost reagent VB100). Forty-eight hours after transfection, the viral supernatant was collected, filtered and the Alstem precipitation solution was added, mixed, and refrigerated at $4^{\circ}\mathrm{C}$ for 4 h, concentrated by centrifugation, and the viral pellet was then resuspended at $100\times$ in cold PBS following the manufacturer's protocol (lentivirus precipitation solution VC100).

T cells for viral infection were activated similarly to non-virally edited cells. Both retroviral and lentiviral transductions occurred 48 h after TCR/cytokine stimulus, followed by expansion in IL-2 similarly to non-virally edited cells. For retroviral transduction, T cells were infected by spinoculation in retronectin-coated (Clontech) plates. Control mock-transduced T cells were also generated. For lentiviral transduction, viral concentrate was added to $1\times$ final concentration.

Antigen-specific TCR expression analysis. The expression of the NY-ESO-1 TCR was assessed in virally and non-virally modified cells with an NY-ESO-1-specific (SLLMWITQC) dextramer-PE (Immundex) according to the manufacturer's protocol. Negative dextramer (Immudex) was used as a negative control.

T cell activation and cytokine production analysis. Melanoma cell lines were established from the biopsies of melanoma patients under the UCLA IRB approval 11-003254. Cell lines were periodically screened for mycoplasma contamination as well as authenticated using GenePrint 10 System (Promega), and were matched with the earliest passage cell lines. M257 (NY-ESO-1+ HLA-A*0201-), M257-A2

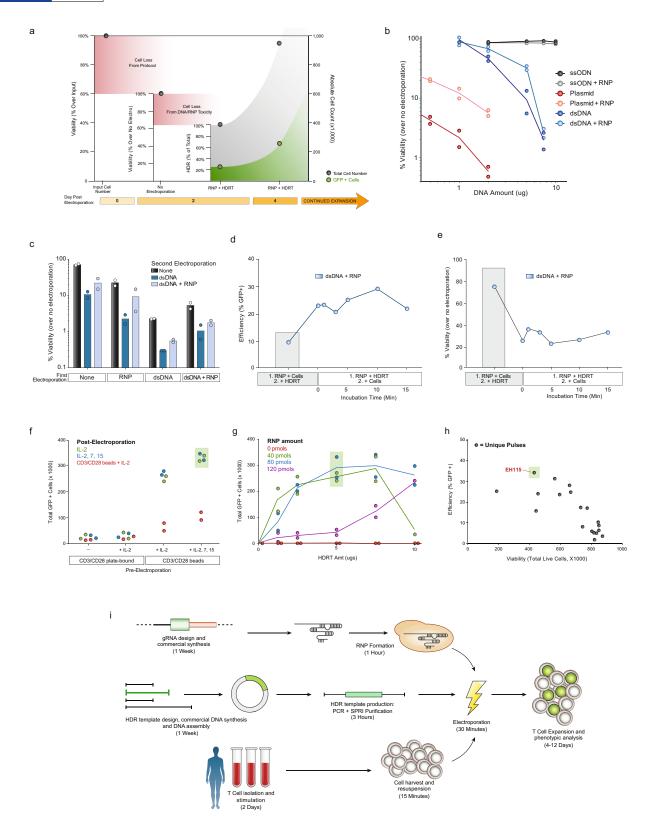
(NY-ESO-1⁺ HLA-A*0201⁺) and M407 (NY-ESO-1⁺ HLA-A*0201⁺) were cocultured 1:1 with the modified PBMCs in cytokine free media. The recommended amount per test of CD107a-APC-H7 (Supplementary Table 2) antibody was added to the co-culture. After 1 h, half the recommended amount of BD Golgi Plug and BD Golgi Stop (BD Bioscience) was added to the coculture. After 6 h, surface staining was performed followed by cell permeabilization using BD cytofix/cytoperm (BD Bioscience) and intracellular staining according to manufacturer instructions (Supplementary Table 2). Negative dextramer and fluorescence minus one staining were used as controls.

T cell in vitro killing assay. M202-nRFP (NY-ESO-1 $^-$, HLA-A*0201 $^+$), M257-nRFP (NY-ESO-1 $^+$ HLA-A*0201 $^-$), M257-A2-nRFP (NY-ESO-1 $^+$ HLA-A*0201 $^+$), M407-nRFP (NY-ESO-1 $^+$ HLA-A*0201 $^+$), and A375-nRFP (NY-ESO-1 $^+$ HLA-A*0201 $^+$) melanoma cell lines stably transduced to express nuclear RFP (Zaretsky 2016 NEJM) were seeded approximately 16 h before starting the co-culture (\sim 1,500 cells seeded per well). Modified T cells were added at the indicated E:T ratios. All experiments were performed in cytokine free media. Cell proliferation and cell death was measured by nRFP real time imaging using an IncuCyte ZOOM (Essen) for 5 days.

In vivo mouse solid tumour model. All mouse experiments were completed under a UCSF Institutional Animal Care and Use Committee protocol. We used 8-12-week-old NSG male mice (Jackson Laboratory) for all experiments. Mice were seeded with tumours by subcutaneous injection into a shaved right flank of 1×10^6 A375 human melanoma cells (ATCC CRL-1619). At seven days after tumour seeding, tumour size was assessed and mice with tumour volumes between 15–30 mm³ were randomly assigned to experimental and control treatment groups. Indicated numbers of T cells were resuspended in 100 µl of serum-free RPMI and injected retro-orbitally. For tumour sizing experiments, the length and width of the tumour was measured using electronic calipers and volume was calculated as $v = 1/6 \times \pi \times \text{length} \times \text{width} \times (\text{length} + \text{width})/2$. The investigator was blinded to experimental treatment group during sizing measurements. A bulk edited T cell population (5 \times 10⁶) or a sorted NY-ESO-1 TCR⁺ population (3 \times 10⁶) was transferred as indicated in figures and legends. For bulk edited T cell transfers, lentivirally edited cells generally had a higher percentage of NY-ESO-1TCR+ cells, so mock-infected cells were added to normalize the percentage of total T cells NY-ESO-1 TCR⁺ to equal that of the bulk population of non-virally edited T cells (\sim 10% NY-ESO-1 TCR $^+$). For sorted T cell transfers, NY-ESO-1 TCR $^+$ T cells were FACS sorted eight days after electroporation, expanded for two additional days, and frozen (Bambanker freezing medium, Bulldog Bio). Non-virally or lentivirally modified human T cells were then thawed and rested in media overnight before adoptive transfer. For flow cytometric analysis of adoptively transferred T cells, single-cell suspensions from tumours and spleens were produced by mechanical dissociation of the tissue through a 70-µm filter. All animal experiments were performed in compliance with relevant ethical regulations per an approved IACUC protocol (UCSF), including a tumour size limit of 2.0 cm in any dimension.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

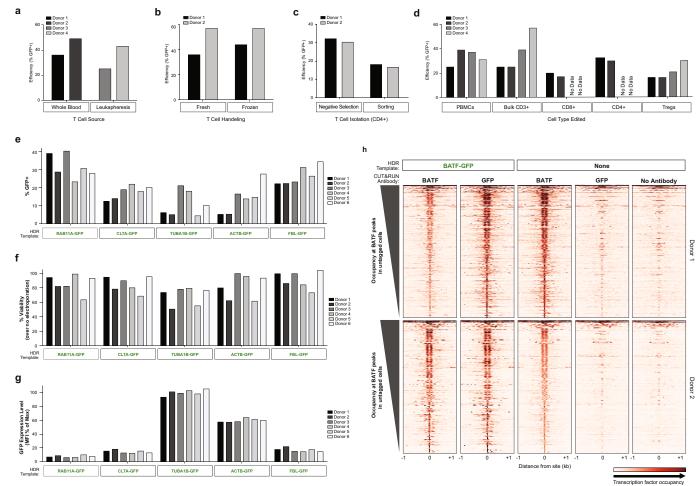
Data availability. CUT&RUN data have been deposited in the Gene Expression Omnibus (GEO) under accession GSE108600. TLA and amplicon sequencing data are available upon request. Source data for animal experiments (Fig. 4g, h and Extended Data Fig. 10) are provided. Plasmids containing the HDR template sequences used in the study are available through AddGene. All other data are available from the corresponding author upon reasonable request.



Extended Data Fig. 1 | See next page for caption.

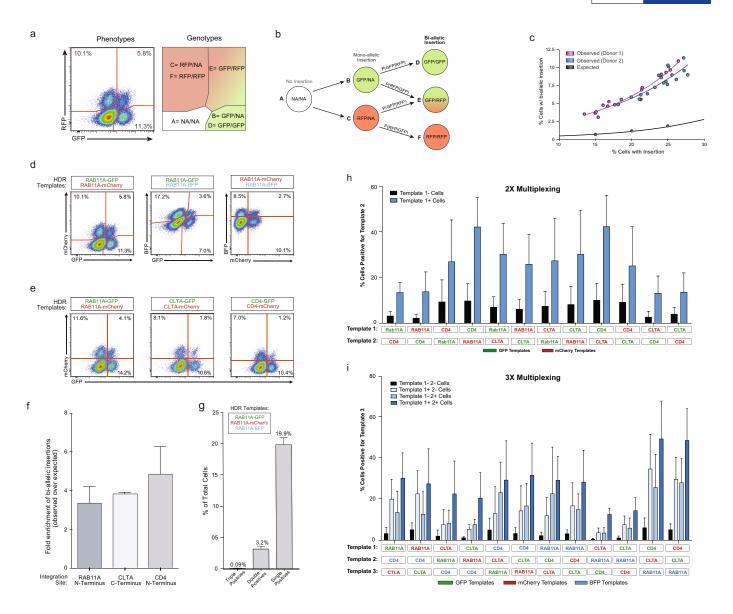
Extended Data Fig. 1 | Development of non-viral genome targeting in primary human T cells. a, Except where noted otherwise, 'viability refers to the number of live cells in an experimental condition (expressed as a percentage) relative to an equivalent population that went through all protocol steps except for the actual electroporation (no electroporation control). 'Efficiency' refers to the percentage of live cells in a culture expressing the 'knocked-in' exogenous sequence (such as GFP). Finally, the total number of cells positive for the desired modification was calculated by multiplying the efficiency by the absolute cell count. Methodological changes that maximized efficiency were not always optimal for the total number of positive cells, and vice-versa. **b**, dsDNA, both circular (plasmid) and linear, when electroporated into primary human T cells, caused marked loss in viability with increasing amounts of template. Co-delivery of an RNP caused less reduction in viability post electroporation. Notably, no loss in viability was seen with ssODNs. c, RNPs must be delivered concurrently with DNA to see increased viability. T cells from two donors were each electroporated twice with an 8 h rest in between electroporations. Although two closely interspersed electroporations caused a high degree of cell death, delivery of the RNP and linear dsDNA template could be delivered separately. Initial RNP electroporation did not protect from the loss of viability if dsDNA was delivered alone in the second round of electroporation. **d**, We determined whether the order of adding reagents influenced targeting efficiency and viability. In wells in which the RNP and the DNA HDR template were mixed together before adding the cells (1. RNP + HDRT; 2. + Cells), there was a marked increase in targeting efficiency. **e**, Note, with the high concentration of dsDNA used in these experiments, viability was higher if the RNP and cells were mixed first and the DNA template was added immediately before electroporation (1. RNP + Cells; 2. + HDRT). Taken together, these data suggest that pre-incubation of the

RNP and HDR template, even for a short period, increased the amount of DNA HDR template delivered into the cell, which increased efficiency but decreased viability. However, viability after RNP and dsDNA HDR template pre-incubation was still higher than was observed with dsDNA HDR template electroporation by itself (**b**). dsDNA HDR temple (5 μg) was used in c-e. f, Primary human T cells were cultured for 2 days using varying combinations of anti-CD3/CD28 TCR stimulation and cytokines before electroporation of RAB11A targeting RNP and HDR template, followed by varying culture conditions after electroporation. g, Among the RNP and HDR template concentrations tested here, optimal GFP insertion into RAB11A was achieved at intermediate concentrations of the RNP and dsDNA HDRT. h, Arrayed testing of electroporation pulse conditions showed that, in general, conditions yielding higher HDR efficiency decreased viability. EH115 was selected to optimize efficiency, while still maintaining sufficient viability. i, Diagrammatic timeline of non-viral genome targeting. Approximately one week is required to design, order from commercial suppliers, and assemble any novel combination of genomic-editing reagents (gRNA and the HDR template). Two days before electroporation, primary human T cells isolated from blood or other sources (Extended Data Fig. 2) are stimulated. dsDNA HDR templates can be made easily by PCR followed by a SPRI purification to achieve a highly concentrated and pure product suitable for electroporation. On the day of electroporation, the gRNA (complexed with Cas9 to form an RNP), the HDR template, and collected stimulated T cells are mixed and electroporated, a process taking approximately 1.5 h. After electroporation, engineered T cells can be readily expanded for an additional 1-2 weeks. Viability was measured 2 days after electroporation and GFP expression was measured at day 4. Graphs display mean (b, c, g, h) and/or individual donor values (\mathbf{b} - \mathbf{h}) in n = 2 independent healthy donors (b-h). For d, e and h, one representative donor is shown.



Extended Data Fig. 2 | Non-viral genome targeting is consistent across T cell types and reproducible across target loci. a, Efficient genome targeting was accomplished with a variety of T cell processing and handling conditions that are used with current manufacturing protocols for cell therapies. Non-viral genome targeting of a RAB11A-GFP fusion protein using a linear dsDNA HDR template was performed in bulk CD3⁺ T cells isolated from either whole blood draws or by leukapheresis. **b**, Targeting was similar either using bulk CD3⁺ T cells fresh after isolation or after cryopreservation (stored in liquid nitrogen and thawed before initial activation). c, CD4⁺ T cells isolated by FACS showed detectable GFP+ cells indicative of efficient editing, albeit at lower rates than targeting in CD4⁺ cells isolated by negative selection (potentially due to the added cellular stress of sorting). d, Using the same optimized non-viral genome targeting protocol (Methods), a variety of T cell types could be successfully edited, including peripheral blood mononuclear cells, without any selection (T cell culture conditions cause preferential growth of T cells from PBMCs). Sorted T cell subsets (CD8⁺, CD4⁺, and $\text{CD4}^{+}\text{IL-2R}\alpha^{+}\text{CD127}^{\text{lo}}$ T_{reg} cells) could be successfully targeted with GFP integration. PBMCs were cultured for 2 days identically to primary T cells (Methods). Bulk CD3⁺ T cells were isolated by negative enrichment. The electroporations in **d** used only 2 μ g of dsDNA HDR template, a concentration that was later found to be less efficient than the final 4 µg (contributing to the lower efficiencies seen compared to Fig. 1d). RAB11A-GFP template was used with on-target gRNA was used in **a**-**d**. e, Four days after electroporation of different GFP templates along with a corresponding RNP into primary CD3⁺ T cells from six healthy donors, GFP expression was observed across both templates and donors. f, High viability after electroporation was similarly seen across target loci. g, The

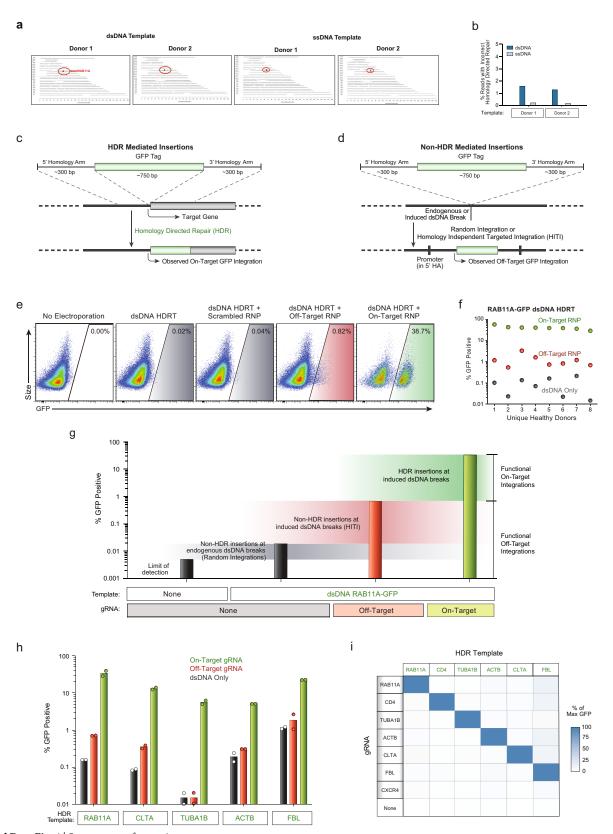
fusion tagged proteins produced by integrating GFP into specific genes localized to the subcellular location of their target protein (Fig. 2b), and were also expressed under the endogenous gene regulation, allowing protein expression levels to be observed in living primary human T cells. Note how GFP tags of the highly expressed cytoskeletal proteins TUBA1B (beta tubulin) and ACTB (beta actin) showed consistently higher levels of expression compared to the other loci targeted across six donors. GFP mean fluorescent intensity (MFI) was calculated for the GFP⁺ cells in each condition/donor, and normalized as a percentage of the maximum GFP MFI observed. h, Gene fusions not only permitted the imaging and analysis of expression of endogenous proteins in live cells, but also could be used for biochemical targeting of specific proteins. For example, chromatin-immunoprecipitation followed by sequencing (ChIP-seq), and more recently CUT&RUN, have been widely used to map transcription factor-binding sites; however, these assays are often limited by the availability of effective and specific antibodies. As a proofof-principle, we used anti-GFP antibodies to perform CUT&RUN analysis in primary T cells in which the endogenous gene encoding the crucial transcription factor BATF had been targeted to generate a GFP-fusion. Binding sites identified with anti-GFP CUT&RUN closely matched the sites identified with an anti-BATF antibody. Anti-BATF, anti-GFP and no-antibody heat maps of CUT&RUN data obtained from primary human T cell populations electroporated with GFP-BATF fusion HDR template (untagged cells were not electroporated). Aligned CUT&RUN binding profiles for each sample were centred on BATF CUT&RUN peaks in untagged cells and ordered by BATF peak intensity in untagged cells. Experiment in h was performed in two independent healthy donors.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Bi-allelic and multiplexed non-viral genome targeting. a, We wanted to confirm that we could generate cells with genome insertions in both alleles and quantify the frequency of bi-allelic modifications. Targeting the two alleles of the same gene with two distinct fluorophores would provide a way to quantify and enrich cells with bi-allelic gene modifications. The possible cellular phenotypes and genotypes when two fluorescent proteins are inserted into the same locus are displayed. Importantly, the number of cells that express both fluorescent proteins underestimates the percentage of cells with biallelic integrations because some cells will have inserted either GFP or mCherry on both alleles. We constructed a model to account for bi-allelic integrations of the same fluorescent protein (Supplementary Note 1). b, Diagram of bi-allelic integration model. The total percentage of cells with bi-allelic HDR integrations must be the sum of genotypes D, E and F. Although the proportion of cells with genotype E (dual fluor positives) is immediately apparent from the phenotypes, genotypes D and F are not. Our model allow for the de-convolution of the multiple genotypes in the single fluor positive phenotypes, and thus an estimation of the true percentage of cells bi-allelic for HDR. c, The observed level of bi-allelic integrations was higher in cells that acquired at least one integration than would be expected by chance. Individual points represent replicates where the combination of the genes encoding the fluorescent proteins was varied (either GFP plus mCherry, GFP plus BFP, or mCherry plus BFP) as was the amount of the HDR template (3–6 μ g). **d**, Bi-allelic HDR analysis was applied across a variety of fluorophore permutations inserted into the RAB11A locus. e, Dual fluorescence bi-allelic integrations were seen across target loci. f, The data also suggest that cells with one integration were more likely to have also undergone a second targeted bi-allelic integration, and this effect was observed across three genomic loci. While the total percentage of cells with an insertion varied with the efficiency of each target site, the fold enrichment in the observed percentage of homozygous cells over that predicted by random chance was largely consistent across loci. g, Co-delivery of three fluorescent tags targeting the RAB11A locus resulted in only a few cells that expressed all three fluorophores, consistent with a low rate of off-target integrations. As a maximum of two

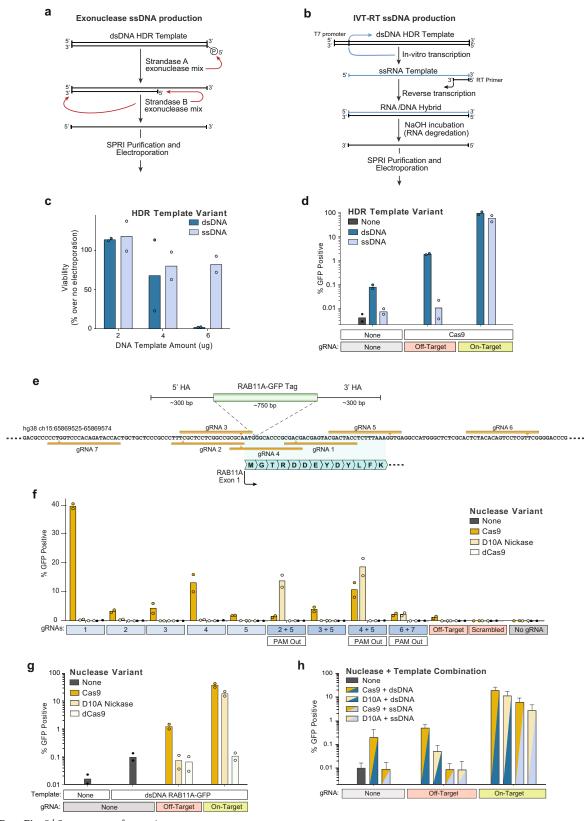
targeted insertions are possible (at the two alleles of the locus; assuming a diploid genome), no cells positive for all three loci should be observed (triple positives). Indeed, while large numbers of single fluorophore integrations were observed (single positives), as well as cells positive for the various permutations of two fluorophores (double positives), there was an approximately 30-fold reduction in the number of triple positive cells compared to double positives. All flow cytometric analysis of fluorescent protein expression shown here was performed 4 days after electroporation. h, Multiplex editing of combinatorial sets of genomic sites would support expanded research and therapeutic applications. We tested whether multiple HDR templates could be co-delivered along with multiple RNPs to generate primary cells in which more than one locus was modified. Primary human T cells with two modifications were enriched by gating on the cells that had at least one modification, and this effect was consistent across multiple combinations of genomic loci. HDR template permutations from a set of six dsDNA HDR templates (targeting RAB11A, CD4 and CLTA; each site with GFP or RFP) were electroporated into CD3⁺ T cells isolated from healthy human donors. Four days after electroporation of the two indicated HDR templates along with their two respective on-target RNPs, the percentage of cells positive for each template was analysed by gating on cells either positive or negative for the other template. Not only was two-template multiplexing possible across a variety of template combinations, but gating on cells positive for one template (template 1+ cells, blue) yielded an enriched population of cells more likely to be positive for the second template compared to cells negative for the first (template 1 – cells, black). 2 μg of each template, along with 30 pmol of each associated RNP, were electroporated for dual multiplexing experiments. i, We also achieved triple gene targeting and could enrich for cells that had a third modification by gating on the cells with two targeted insertions, an effect again consistent across target genomic loci. 1.5 μg of each template (4.5 μg total) were electroporated together with 20 pmol of each corresponding RNP (60 pmol total). Graphs display mean and s.d. in n = 4 (f-i) independent healthy donors. Other experiments (c-e) were performed in two independent healthy donors.



Extended Data Fig. 4 \mid See next page for caption.

Extended Data Fig. 4 | Examination of off-target integrations with non**viral genome targeting. a**, Results of targeted locus amplification (TLA) sequencing. No off-target integration sites were identified (assay's limit of detection \sim 1% of alleles) with either a dsDNA or ssDNA HDR template in two healthy donors. The on-target RAB11A locus on chromosome 15 is indicated in red. b, The frequency of one of the observed incorrect integrations at the target locus was reduced using a long ssDNA HDR template in two human blood donors (Supplementary Note 2). c, Diagram of HDR-mediated insertions at the N terminus of a target locus. The homology arms specify the exact sequence where the insert (a GFP tag in this case) will be inserted, allowing for scarless integration of exogenous sequences. Because a GFP fusion protein is created, GFP fluorescence will be seen as a result of the on-target integration, which is dependent on an RNP cutting adjacent to the integration site. **d**, dsDNA can be integrated via homology-independent repair mechanisms at off-target sites through either random integration at naturally occurring dsDNA breaks, or potentially at induced double-stranded breaks, such as those at the off-target cut sites of the RNP. This effect can be harnessed to allow for targeted integration of a dsDNA sequence at a desired induced dsDNA break in quiescent cell types which lack the ability to do HDR, but crucially the entire sequence of the dsDNA template is integrated, including any homology arms. In the case that the homology arms contain a promoter sequence (such as for N-terminal fusion tags), these off target integrations can drive observable expression of the inserted sequence without the desired correct HDR insertion. e, We looked for unintended non-homologous integrations with the non-viral system using an N-terminal GFP-RAB11A fusion construct that contained the endogenous RAB11A promoter sequence within its 5' homology arm. This construct could express GFP at off-target integration sites, which allowed us to assay for off-target events at the single-cell level using flow cytometry. Inclusion of a gRNA designed to cut a genome region that is not the homologous region to the targeting sequence can be used to infer integration at an off-target cut site. f, Although efficient GFP expression

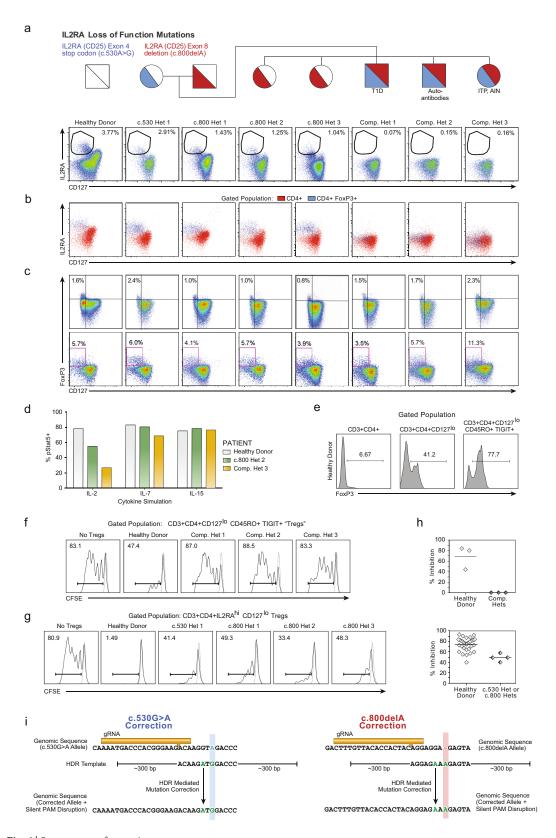
depended on pairing the HDR template with the correct gRNA targeting that site, rare GFP⁺ cells were observed when dsDNA HDR templates were delivered either alone (\sim 0.1%) or with an off-target Cas9 RNP (\sim 1%). **g**, Quantification of different types of functional off-target integrations. The increase in the percentage of fluorescent cells over the limit of detection when the template alone is electroporated probably represents random integrations at naturally occurring dsDNA breaks (although cutindependent integration at the homology site is also possible in theory). Not every off-target integration will yield fluorescent protein expression (for example, only part of the template sequence could be integrated or it could be integrated in a way that does not lead to measurable expression), but the relative differences in functional off-target expression between different templates and editing conditions can be assayed. Inclusion of an RNP targeting CXCR4 (off-target) markedly increased the observed off-target homology-independent integrations, probably by a homology-independent insertion event. As expected, efficient GFP expression as expected was only seen with the correct gRNA sequence and HDR-mediated repair. Bars represent observed GFP⁺ percentages from T cells from one representative donor electroporated with the indicated components. h, Comparisons of on-target GFP expression versus functional off-target integrations across five templates reveal HDR is highly specific, but that off-target integrations can be observed at low frequencies. i, A matrix of gRNAs and HDR templates were electroporated into bulk T cells from two healthy donors. The average GFP expression in gated CD4⁺ T cells as a percentage of the maximum observed for a given template is displayed. Across six unique HDR templates and gRNAs, ontarget HDR-mediated integration was the by far most efficient. One HDR template, a C-terminal GFP fusion tag into the nuclear factor FBL, had consistently higher off-target expression across gRNAs, potentially due to a gene-trap effect as the 3' homology arm for FBL contains a splice-site acceptor followed by the final exon of FBL leading into the GFP fusion. n = 2 (**a**, **b**, **h**, **i**) or n = 8 (**e**, **f**) independent healthy donors.



Extended Data Fig. 5 \mid See next page for caption.

Extended Data Fig. 5 | Non-viral genome targeting using long ssDNA HDR templates and a Cas9 nickase. a, Long ssDNA templates have potential to reduce homology-independent integrations while preserving on-target efficiency. One method to generate long ssDNA templates involves a two-step selective exonuclease digestion that specifically degrades one strand of a PCR product that has been labelled by 5' phosphorylation, which can be easily added to a PCR primer before amplification. b, We also applied a second ssDNA production method based on sequential in vitro transcription (IVT) and reverse transcription (RT) reaction. A PCR product with a short T7 promoter appended serves as an IVT template to produce a ssRNA product. After annealing of an RT primer and reverse transcription, an RNA-DNA hybrid can form, which then can be transformed into a long ssDNA template by incubation in sodium hydroxide, which selectively degrades the RNA strand. c, At 4 days after electroporation, varying concentrations of a long ssDNA HDR templates (\sim 1.3 kb) did not show the decreased viability observed in CD3⁺ T cells electroporated with a linear dsDNA HDR template of the same length. d, Electroporation of a ssDNA HDR template reduced offtarget integrations to the limit of detection (that is, comparable to levels seen with no template electroporated) both with no nuclease added and at induced off-target dsDNA breaks (off-target gRNA + Cas9). e, Diagram of the genomic locus containing the first exon of RAB11A. Use of spCas9 with an individual guide RNA (gRNA 1, 'on-target' in d) along with a dsDNA HDR template integrating a GFP in frame with RAB11A directly

after the start codon results in efficient GFP expression (Fig. 1d). Use of a Cas9 nickase (D10A variant) with two gRNAs may reduce the incidence of off-target genome cutting. f, A series of individual gRNAs as well as dual gRNA combinations were tested for GFP insertion efficiency at the RAB11A N-terminal locus. As expected, no gRNAs showed appreciable levels of GFP insertion when using a nuclease dead Cas9 (dCas9). Multiple individual gRNAs that cut adjacent to the insertion site showed GFP integration when used with Cas9, but none were as efficient as gRNA 1. The D10A nickase showed little to no GFP integration with individual guides, but multiple two-guide combinations showed efficient GFP integration. Only in gRNA combinations where the two PAM sequences were directed away from each other (PAM Out) was GFP integration seen. g, GFP integration efficiencies as presented in f but graphed on a logarithmic scale reveal lower levels of functional off-target integrations when using the D10A nickase compared to spCas9 (with an individual off-target gRNA, targeting CXCR4), probably due to the requirement for the D10A nickase to have two gRNAs bound in close proximity to induce a dsDNA break. h, Long ssDNA templates (~1.3 kb) could be successfully combined with Cas9 nickases (D10A) for targeted integration, similar to linear dsDNA templates. Here, long ssDNA HDR templates with D10A nickase showed lower efficiencies of GFP integration at the RAB11A site. n = 2 (**c**, **d**, **f**, **g**) or n = 3 (**h**) independent healthy donors with mean (c, d, f-h) and s.d. (h).



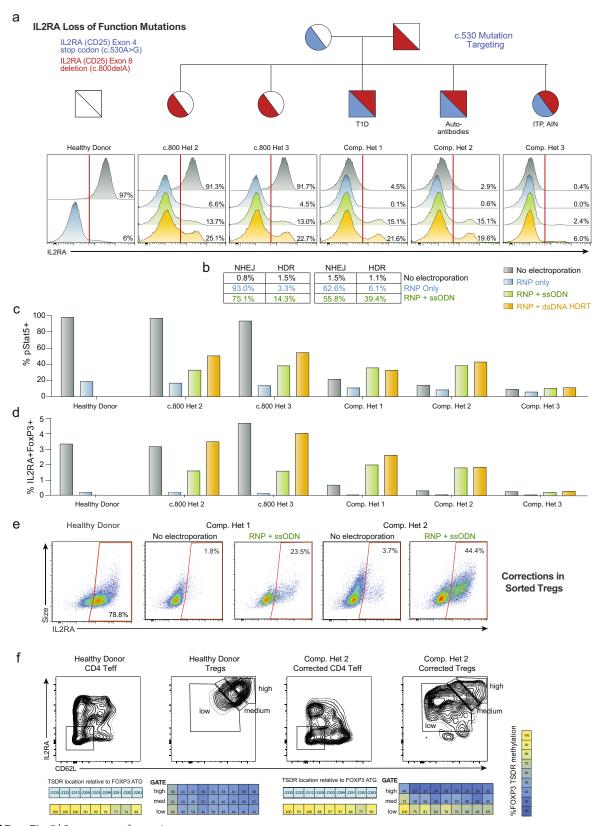
Extended Data Fig. 6 \mid See next page for caption.



Extended Data Fig. 6 | Reduced T_{reg} cell frequencies and function in subjects with two loss-of-function *IL2RA* mutations. a, CD4⁺

T cells from a healthy donor and all family members, including IL2RA heterozygotes (c.530 het 1, c.800 hets 1-3) as well as compound heterozygous children (comp. hets 1–3), with loss-of-function *IL2RA* mutations were analysed by flow cytometry to assess the presence of IL-2R $\alpha^{hi}CD127^{lo}$ T_{reg} cells. $\boldsymbol{b},$ In healthy donors and individuals with only one IL2RA mutation, CD4⁺FOXP3⁺ T cells are predominantly IL- $2R\alpha^{hi}CD127^{lo}$. In the compound heterozygotes, a CD127^{lo}CD4⁺FOXP3⁺ population is present, but does not express high levels of IL-2R α . c, Clinical phenotyping performed at two separate sites showed that compound heterozygotes have CD127loFOXP3+ cells. d, Deficiency in IL-2R α surface expression in compound heterozygote 3 led to aberrant downstream signalling as measured by pSTAT5 expression after stimulation with IL-2, but not IL-7 or IL-15. e, Owing to the inability to sort IL-2R α^{hi} T_{reg} cells from the IL-2R α -deficient compound heterozygotes, FOXP3⁺ cells were enriched from CD4⁺ using an alternate gating strategy that used the surface markers CD127loCD45RO+TIGIT+. Intracellular FOXP3 staining of T cells from the indicated gated population is shown. **f**, Although these CD3⁺CD4⁺CD127^{lo}CD45RO⁺TIGIT potential T_{reg} cells were highly enriched for FOXP3 and showed some suppressive capacity when cultured with CFSE-labelled stimulated T_{resp} cells from healthy donors, CD3⁺CD4⁺CD127^{lo}CD45RO⁺TIGIT⁺ from the compound heterozygotes did not show suppressive ability. Stimulated T_{resp} cell population (solid curves), non-stimulated T_{resp} cells (dashed curve). g, Correction of either IL2RA mutation in the compound heterozygotes individually would still leave the other mutation, leaving the cells as single heterozygotes. To confirm that such a potential correction would result in some level of functional suppression, we assessed the suppressive ability of CD4 $^+$ IL-2R α^{hi} CD127 lo T_{reg} cells

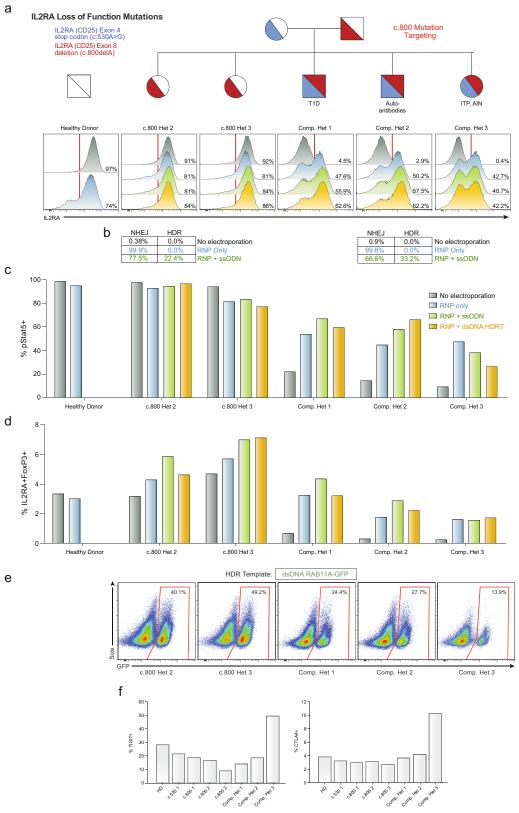
from the c.530 and c.800 single heterozygote family members as in f. h, Dot plot summaries of T_{reg} cell suppressive ability in cells from healthy donors (n = 3 with single (top) or 12 (bottom) technical replicates), *IL2RA*-deficient compound heterozygotes (\mathbf{f} , n = 3 total human subjects) and IL2RA +/- c.530 or c.800 heterozygotes (g, n=4 total human subjects). Although CD3⁺CD4⁺CD127^{lo}CD45RO⁺TIGIT⁺ T_{reg} cells from compound heterozygotes showed no suppressive ability, conventional $\text{CD4}^{+}\text{IL-2R}\alpha^{\text{hi}}\text{CD127}^{\text{lo}}\,T_{\text{reg}}$ cells from the single heterozygote family members showed some suppressive capacity, consistent with their lack of a pronounced clinical phenotype compared to the compound heterozygotes. Thus, correcting functional IL-2Rα expression on the surface of FOXP3⁺ T cells from these patients may represent a viable approach for developing an ex vivo gene therapy. Mean value is displayed. i, Initial genetic testing of the proband (Supplementary Note 3) using an in-house targeted nextgeneration sequencing multi-gene panel of over 40 genes known to be involved in monogenic forms of diabetes was negative. Subsequent exome sequencing in the trio of proband and parents revealed two causative mutations in the *IL2RA* gene. The mother possessed a single heterozygous mutation (c.530G>A) in exon 4 of *IL2RA*, resulting in a premature stop codon. The father possessed a single heterozygous mutation (c.800delA) in exon 8 of IL2RA, resulting in a frameshift mutation leading to a 95 amino acid long run-on. Sanger sequencing confirmed that the proband was a compound heterozygote with both mutations. A gRNA was designed to cut adjacent to the site of each mutation, 8 bp away for c.530 mutation (blue), and 7 bp away for c.800 (red). For each mutation, an HDR template was designed including the corrected sequence (green) as well as a silent mutation in a degenerate base to disrupt the PAM sequence (NGG) for each guide RNA. Displayed genomic regions (not to scale) for c.530 mutation site (hg38 ch10:6021526-6021557) and c800 mutation site (hg38 ch10:6012886-6012917).



Extended Data Fig. 7 \mid See next page for caption.

Extended Data Fig. 7 | HDR-mediated correction of IL2RA c.530A>G **loss-of-function mutation. a**, Unlike the gRNA targeting the c.800delA mutation at the C terminus of IL-2R α (Extended Data Fig. 8), the gRNA targeting the c.530A>G mutation (causing a stop codon in an interior exon) results in substantial (\sim 90%) loss of IL-2R α cell surface expression in a healthy donor and single heterozygotes (c.800 het 2 and 3) 2 days after electroporation of the RNP alone (blue) into CD3⁺ T cells. Although starting from a very small IL-2R α ⁺ percentage, this reduction was observed in all three compound heterozygotes, potentially because a small amount of protein can be surface expressed from the c.800delA allele. This reduced IL-2R α expression could be partially rescued by inclusion of an ssODN HDR template (green) and even more substantially rescued using a large dsDNA HDR template (yellow). Both template types contained the corrected sequence, a silent mutation to remove the gRNA PAM sequence, and either 60 bp (ssODNs) or ~300 bp (large dsDNA) homology arms (Extended Data Fig. 6i). b, Amplicon sequencing of the c.530 site in select patients shows the correlation between IL-2R α cell surface expression and genomic correction. Small numbers of reads in the 'no electroporation' and 'RNP only' conditions were called as HDR, potentially owing to small amounts of cross-well contamination. c, Increased pSTAT5 in response to IL-2 stimulation (200 U ml⁻¹) 7 days after electroporation in CD3⁺ T cells from compound heterozygote patients undergoing HDR-mediated mutation correction compared to no electroporation or RNP only controls. pSTAT5+ cells correlated with increased IL-2R α surface expression. **d**, Similarly, increased proportions of IL-2Rα⁺FOXP3⁺ cells are seen 9 days after electroporation in the HDR correction conditions in compound heterozygote patients. Lower percentages of correction were seen when targeting the c.530 mutation

for HDR correction in compound heterozygote 3, potentially due altered cell-state associated with the patient's disease or the patient's immunosuppressive drug regimen (Supplementary Table 4). e, Mutation correction was possible in sorted T_{reg}-like cells from the affected patients. ${\rm CD3^+CD4^+CD127^{lo}CD45RO^+TIGIT^+} \ T_{\rm reg} \ cells, a \ population \ highly$ enriched for FOXP3⁺ cells (Extended Data Fig. 6e), identified without the traditional T_{reg} cell IL-2R α surface marker (absent due to the causative mutations), were FACS-sorted and underwent correction of the c.530A>G mutation using a Cas9 nuclease and short ssDNA HDR template (ssODN). After 12 days in culture, during which time the cells expanded more than 100-fold, greater than 20% (compound het 1) and 40% (compound het 2) of targeted cells expressed IL-2R α on their surface, demonstrating functional correction and expansion of a therapeutically relevant cell type. In these experiments, expansion was less robust for cells from compound het 3. f, After 12 days in culture, corrected Treeg cells from compound heterozygote 2, and a female healthy control, were sorted based on IL-2R α and CD62L expression. Methylation of the TSDR (T_{reg} -cell-specific demethylated region) of FOXP3 intron 1 was analysed in the indicated sorted cell populations by bisulfite sequencing (Epigendx). Owing to X-chromosome inactivation, incomplete demethylation is observed in the control T_{reg} cell populations from the female healthy donor. The sorted IL-2R α^{high} CD62L high population of corrected T_{reg} cell showed increasing TSDR demethylation, whereas similarly edited and expanded CD4⁺ T effector (T_{eff}) cells did not show substantial TSDR demethylation in the healthy donor or in corrected cells from compound heterozygote 2. All electroporations were performed according to optimized non-viral genome targeting protocol (Methods). For ssODN electroporations, 100 pmol in 1 µl water was electroporated.



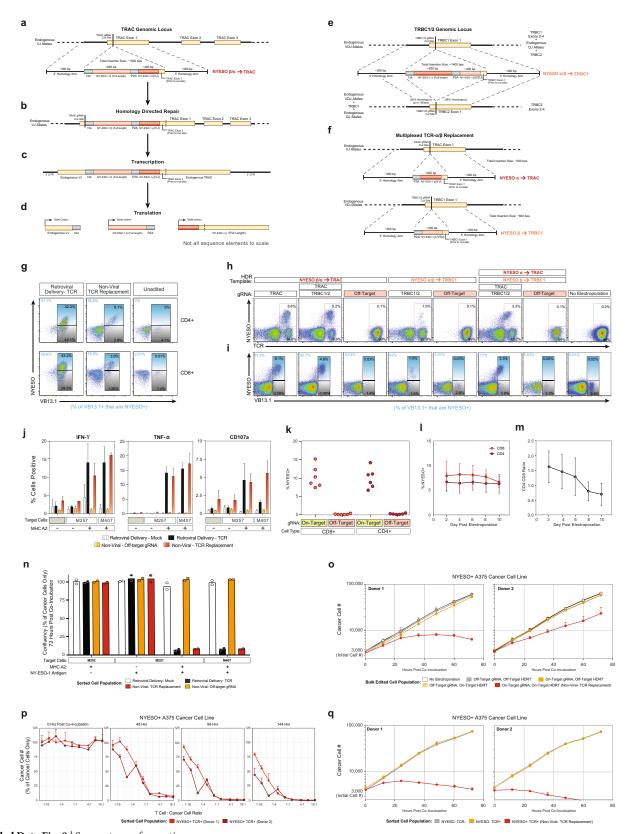
Extended Data Fig. 8 | See next page for caption.



Extended Data Fig. 8 | HDR-mediated and non-HDR-mediated correction of IL2RA c.800delA frameshift loss-of-function mutation.

a, Histograms of IL-2R α surface expression in CD3⁺ T cells in all children from a family carrying two loss-of-function IL2RA mutations, including three compound heterozygotes that express minimal amounts of IL-2R α on the surface of the T cells (no electroporation, grey). Two days after electroporation of an RNP containing a gRNA for the site of one of the two mutations, a 1-bp deletion in the final exon of IL2RA (c.800delA) causing a run-on past the normal stop codon, CD3⁺ T cells from a healthy donor and single heterozygotes (c.800 het 2 and 3) showed slight increases in IL-2R α^- cells (RNP only, blue). This modest change is potentially due to the gRNA targeting the C terminus of the protein, in which small indels may cause less pronounced loss of surface protein expression. Notably, the RNP alone resulted in IL-2R α surface expression in almost 50% of edited T cells in all three compound heterozygotes. In cells from two of the compound heterozygous children, increases in the percentage of cells with IL-2R α correction compared to RNP only could be achieved by inclusion of an ssODN HDR template sequence with the mutation correction (RNP plus ssODN, green), and further increased at this site when using a longer dsDNA HDR template to correct the mutation (RNP plus dsDNA HDRT, yellow) (Extended Data Fig. 6i). b, Amplicon sequencing was performed in select targeted patient cells. c, pSTAT5 in response to high dose IL-2 stimulation (200 U ml⁻¹) in targeted CD3⁺ T cells after 7 days of expansion post-electroporation. Increased numbers of pSTAT5⁺ cells correlated with increased IL-2R α surface expression (a). d, After 9 days of expansion post-electroporation, intracellular FOXP3 staining revealed an increased proportion of IL-2Rα⁺ FOXP3⁺ cells in CD3⁺ T cells compared to no electroporation controls. Electroporations were performed

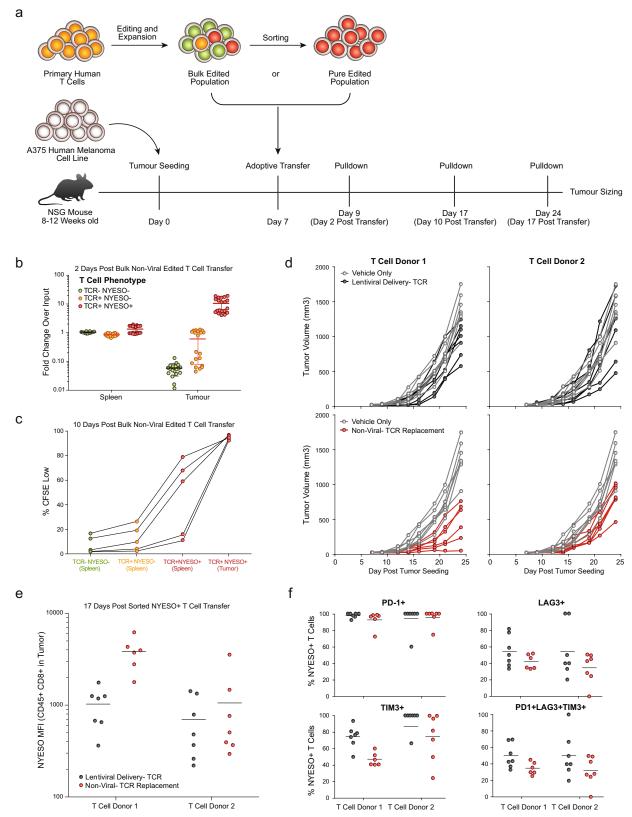
according to optimized non-viral genome targeting protocol (Methods). For ssODN electroporations, 100 pmol in 1 µl water was electroporated. e, Flow cytometric analysis of GFP expression 6 days after electroporation of a positive HDR control RAB11A-GFP dsDNA HDR template into CD3⁺ T cells from the indicated patients revealed lower GFP expression in the three compound heterozygotes compared to their two c.800 heterozygote siblings. Compared to a cohort of 12 similarly edited healthy donors (Fig. 1d), both c.800 heterozygotes as well as compound heterozygotes 1 and 2 were within the general range observed across healthy donors, whereas compound heterozygote 3 had lower GFP expression than any healthy donor analysed. Of note, in compound heterozygote 3, HDR-mediated correction at the c.530 mutation was substantially lower than the other two compound heterozygotes (Fig. 3b). IL-2R α surface expression after electroporation of the c.800delA targeting RNP alone was similar though. Compared to HDR-mediated repair, NHEJ-mediated frameshift correction at c.800delA may be less dependent on cell proliferation, consistent with compound heterozygote 3 being the only compound heterozygous patient on active immunosuppressants at the time of blood draw and T cell isolation (Supplementary Note 3). f, Altered cell-state associated with the patient's disease could also contribute to diminished HDR rates. TIGIT and CTLA4 expression levels in non-edited, isolated CD4 $^{+}$ T cells from each indicated patient was measured by flow cytometry. Consistent with altered cell states and or/cell populations, cells from compound heterozygote 3 had a distinct phenotype, with increased TIGIT and CTLA4 expression compared both to healthy donors, the single heterozygous family members, as well the other two compound heterozygous siblings.



Extended Data Fig. 9 \mid See next page for caption.

Extended Data Fig. 9 | Endogenous TCR replacement strategy and **functional characterization. a–d**, Schematic description of HDR template for endogenous TCR replacement by in-frame integration of a new TCR-β chain and a new variable region of a TCR- α chain at the TCR- α locus, and subsequent transcription and translation of the new TCR. e, HDR template for endogenous TCR replacement at the TCR-β locus. f, Multiplexed integration of a new TCR- α at the TCR- α locus and a new TCR- β at the TCR-β locus. See Supplementary Note 4 for detailed description of TCR replacement strategy. g, TCR mispair analysis after retroviral delivery or non-viral TCR replacement of an NY-ESO-1-specific TCR in gated CD4+ or CD8⁺ T cells. With viral introduction of the new TCR, an infected cell will potentially express at least four different TCRs (new TCR- α plus new TCR- β ; new TCR- α plus endogenous TCR- β ; endogenous TCR- α and new TCR- β ; endogenous TCR- α plus endogenous TCR- β). Staining for the specific beta chain in the new introduced TCR (VB13.1) along with MHC-peptide multimer (NY-ESO) can provide a rough estimate of TCR mispairing by distinguishing between cells that predominantly expressed the introduced TCR (VB13.1⁺ NY-ESO⁺; new TCR- α and new TCR- β) versus those that expressed predominantly one of the potential mispaired TCRs (VB13.1⁺ NY-ESO⁻; endogenous TCR- α and new TCR- β). **h**, **i**, TCR replacement by targeting an entire new TCR into TRAC (a-d, also possible with a multiplexed knockout of TCRB), an entire new TCR into TRBC1/2 (f), or multiplexed replacement with a new TCR- α into TRAC and a new TCR- β into *TRBC1/2*. **j**, Functional cytokine production was observed selectively after antigen exposure in gated $\mbox{CD}4^+\mbox{\,T}$ cells, similarly to gated CD8⁺ T cells (Fig. 4c). k, Non-viral TCR replacement was consistently observed at four days after electroporation in both gated $\mathrm{CD8^+}$ and $\mathrm{CD4^+}$ T cells across a cohort of six healthy blood donors. I, In a second cohort of six additional healthy blood donors, 100 million T cells from each donor were electroporated with the NY-ESO-1 TCR replacement HDR template and on-target gRNA/Cas9 (Fig. 4f). The percentage of CD4+ and CD8+ T cells that were NY-ESO-1 TCR⁺ was consistent over 10 days of expansion after electroporation. m, Over 10 days of expansion after non-viral genome targeting, CD8+ T cells showed a slight proliferative advantage over CD4+ T cells. n, The indicated melanoma cell lines were co-incubated with the

indicated sorted T cell populations at a ratio of 1:5 T cells to cancer cells. At 72 h after co-incubation, the percentage cancer cell confluency was recorded with by automated microscopy (in which nuclear RFP marks the cancer cells). T cells expressing the NY-ESO-1 antigen-specific TCR, either by retroviral transduction (black) or by non-viral knock-in endogenous TCR replacement (red) both showed robust target cell killing only in the target cancer cell lines expressing both NY-ESO-1 and the HLA-A*0201 class I MHC allele. o, To ensure that target cell killing by non-viral TCR replacement T cells (red) was not due to either the gRNA or the HDR template used for TCR replacement alone, a matrix of on/off target gRNAs and on/off target HDR templates was assayed for target cell killing of the NY-ESO-1 $^+$ HLA-A*0201 $^+$ A375 cancer cell line (off-target gRNA and HDRT were specific for RAB11A-GFP fusion protein knock-in). Only cells with both the on-target gRNA as well as the on-target HDR template demonstrated target cell killing. p, Sorted NY-ESO-1+ TCR+ cells from a bulk T cell edited population (on-target gRNA, on-target HDR template) showed a strong dose-response effect for target cancer cell killing. Within 48 h, T cell to cancer cell ratios of 2:1 and greater showed almost complete killing of the target cancer cells. By 144 h, T cell to cancer cell ratios of less than 1:16 showed evidence of robust target cell killing. q, Target cell killing by non-viral TCR replacement T cells was due specifically to the NY-ESO-1-recognizing TCR⁺ cell population observed by flow cytometry after non-viral TCR replacement (Fig. 4b). Starting with the bulk edited T cell population (all of which had been electroporated with the on-target gRNA and HDR template), we separately sorted three populations of cells: the NY-ESO-1⁺ TCR⁺ cells (non-virally replaced TCR) (red), the NY-ESO-1 TCR^- cells (TCR-knockout) (grey), and the NY-ESO-1 $^-$ TCR $^+$ cells (those that probably retained their native TCR but did not have the NY-ESO-1-specific knock-in TCR) (orange). Only the sorted NY-ESO-1⁺ TCR⁺ population demonstrated target cell killing (4:1 T cell to cancer cell ratio). One representative donor from n = 2 (**g**, **j**) or n = 3 (**h**, **i**) independent healthy donors with mean and s.d. of technical triplicates (j). Mean and s.d. of n = 6 independent healthy donors (l, m) or of four technical replicates for n = 2 independent healthy donors $(\mathbf{o} - \mathbf{q})$ are shown. Mean and individual values for n = 2 independent healthy donors (**n**).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | In vivo functionality of T cells with non-viral **TCR replacement. a**, Diagram of in vivo human antigen-specific tumour xenograft model. NSG mice (8–12 weeks old) were seeded with 1×10^6 A375 cells (human melanoma cell line; NY-ESO-1 antigen⁺ and HLA-A*0201⁺) subcutaneously in a shaved flank. Primary human T cells edited to express an NY-ESO-1 antigen-specific TCR were generated (either by lentiviral transduction or non-viral TCR replacement), expanded for 10 days after transduction or electroporation, and frozen. Either a bulk-edited population was used (**b**, **c**) or an NY-ESO-1 TCR⁺-sorted population was used (d-f). At 7 days after tumour seeding, T cells were thawed and adoptively transferred via retro-orbital injection. b, Two days after transfer of 5×10^6 bulk non-virally targeted T cells ($\sim 10\%$ TCR⁺ NY-ESO-1⁺ (red), \sim 10% TCR⁺ NY-ESO-1⁻ (orange), and \sim 80% TCR⁻ NY-ESO-1⁻ (green), see Fig. 4b), NY-ESO-1⁺ non-virally edited T cells preferentially accumulated in the tumour versus the spleen. n = 5 mice for each of four human T cell donors. c, Ten days after transfer of 5×10^6 bulk non-virally targeted CFSE-labelled T cells, NY-ESO-1⁺ TCR⁺ cells showed greater proliferation than TCR⁻ or TCR⁺NY-ESO-1⁻ T cells, and showed greater proliferation (CFSE low) in the tumour than in the spleen. Ten days after transfer, TCR⁻ and TCR⁺ NY-ESO-1⁻ T cells were difficult

to find in the tumour (Fig. 4g). d, Individual longitudinal tumour volume tracks for data summarized in Fig. 4h. Sorted NY-ESO-1 TCR+ T cells (3×10^6) generated either by lentiviral transduction (black) or non-viral TCR replacement (red) were transferred on day 7 after tumour seeding and compared to vehicle-only injections until 24 days after tumour seeding. Note that the same data for vehicle control data are shown for each donor in comparison to lentiviral delivery (above) and non-viral TCR replacement (below). e, f, In these experiments, 17 days after T cell transfer (d), non-virally TCR-replaced cells appeared to show greater NY-ESO-1 TCR expression and lower expression of exhaustion markers. Transfer of both lentivirally transduced and non-viral TCR replaced cells showed reductions in tumour burden on day 24. In this experimental model, non-viral TCR replacement showed further reductions compared to the lentiviral transduction (Fig. 4h), potentially due to knockout of the endogenous TCR, endogenous regulation of expression of the new TCR, some difference in the cell populations amenable to non-viral versus lentiviral editing, or confounding variables in cell handling between lentiviral transduction and non-viral genome targeting. n = 4 (**b**), n = 2 $(\mathbf{d}-\mathbf{f})$, or n=1 (c) independent healthy donors in 5 (b, c) or 7 (d-f) mice per donor with mean (b, e, f) and s.d. (b).



Mechanism of parkin activation by PINK1

Christina Gladkova¹, Sarah L. Maslen¹, J. Mark Skehel¹ & David Komander¹*

Mutations in the E3 ubiquitin ligase parkin (PARK2, also known as PRKN) and the protein kinase PINK1 (also known as PARK6) are linked to autosomal-recessive juvenile parkinsonism (AR-JP)^{1,2}; at the cellular level, these mutations cause defects in mitophagy, the process that organizes the destruction of damaged mitochondria^{3,4}. Parkin is autoinhibited, and requires activation by PINK1, which phosphorylates Ser65 in ubiquitin and in the parkin ubiquitin-like (Ubl) domain. Parkin binds phospho-ubiquitin, which enables efficient parkin phosphorylation; however, the enzyme remains autoinhibited with an inaccessible active site^{5,6}. It is unclear how phosphorylation of parkin activates the molecule. Here we follow the activation of full-length human parkin by hydrogendeuterium exchange mass spectrometry, and reveal large-scale domain rearrangement in the activation process, during which the phospho-Ubl rebinds to the parkin core and releases the catalytic RING2 domain. A 1.8 Å crystal structure of phosphorylated human parkin reveals the binding site of the phospho-Ubl on the unique parkin domain (UPD), involving a phosphate-binding pocket lined by AR-JP mutations. Notably, a conserved linker region between Ubl and the UPD acts as an activating element (ACT) that contributes to RING2 release by mimicking RING2 interactions on the UPD, explaining further AR-JP mutations. Our data show how autoinhibition in parkin is resolved, and suggest a mechanism for how parkin ubiquitinates its substrates via an untethered RING2 domain. These findings open new avenues for the design of parkin activators for clinical use.

Work in the past decade has shown how PINK1 and parkin initiate mitophagy, and many steps in this process are mechanistically well understood^{3,4}. It has further been suggested that targeted activation of either PINK1 or parkin could increase mitochondrial turnover and impede the progression of Parkinson's disease. A detailed understanding of the underlying molecular mechanisms of these processes is therefore essential.

Parkin requires an elaborate activation mechanism. The first crystal structures of parkin^{7–9} revealed several distinct mechanisms of autoinhibition (Fig. 1a, Extended Data Fig. 1a). Most strikingly, the active site Cys on the catalytic RING2 domain, which receives ubiquitin from the E2 enzyme, is obstructed by an interface with the UPD (also known as RING0) (Extended Data Fig. 1a). The RING2–UPD interface is highly hydrophobic^{7–9} (Extended Data Fig. 1b), and it is not clear how this intramolecular interaction can be opened.

Activation of parkin is mediated by the mitochondrial outer membrane (MOM) Ser/Thr protein kinase PINK1, which phosphorylates Ser65 in ubiquitin (generating phospho-ubiquitin)^{10–14} and in the parkin Ubl domain^{15–17}. A current model for PINK1-mediated activation of parkin suggests that PINK1 phosphorylates ubiquitin attached to MOM proteins, and autoinhibited, cytosolic parkin is recruited with nanomolar affinity to sites of PINK1 activity^{3,5,13,18–21}. Binding of phospho-ubiquitin induces conformational changes in parkin that lead to the release of the Ubl domain from the parkin core, and enable PINK1 to phosphorylate the parkin Ubl domain^{3,5–7,13,18–22} (Fig. 1a, Extended Data Fig. 1c). Notably, in structures of parkin bound to phospho-ubiquitin^{5,6}, parkin is still autoinhibited; the E2 binding site remains blocked by the repressor (REP) element, and RING2 and

its catalytic Cys remain obstructed by the UPD (Fig. 1a, Extended Data Fig. 1c).

Indeed, full activation of parkin requires phosphorylation of its Ubl. A parkin S65A mutant is not retained at mitochondria, is unable to trigger mitochondrial ubiquitination and mitophagy, and thus is physiologically inactive 13,15,17,21,23 . Biochemically, parkin phosphorylation enhances activity to a greater extent than binding of phosphoubiquitin 13,21,24 , and parkin phosphorylation, but not phosphoubiquitin binding, enables ubiquitin activity-based probes (Ub-ABPs) to access the active site $\mathrm{Cys}^{5,13,25}$. How Ubl phosphorylation is able to activate parkin, and in particular, how it can disrupt the RING2–UPD interface, has remained unknown, and this has led to various models of parkin activation 3,5,6,8,9,26 .

We reconstituted activation of full-length human parkin by PINK1, and followed domain rearrangements by hydrogen-deuterium exchange mass spectrometry (HDX-MS)²⁷ (Fig. 1b-e, Extended Data Figs. 2, 3). HDX-MS reports on the relative rate of exchange of backbone amide hydrogens with deuterium, based on the strength of hydrogen bonding and solvent accessibility in the folded protein, and distinguishes peptides in a protein's core (which show no or little exchange with solvent over time) from those at an exposed surface (which show high or increasing exchange with solvent over time). The power of the method lies in its ability to compare identical peptides between different states along an activation cascade, revealing peptides that become exposed and thus interfaces that are opened (red in Fig. 1b-e), and regions in the protein that become protected and form new interfaces (blue in Fig. 1b-e). For parkin, this allowed us to confirm previously reported conformational changes upon phosphoubiquitin binding^{5,6}, whereby the parkin Ubl is released and becomes exposed to solvent (numbers 1 and 2 in Fig. 1b), the phospho-ubiquitin binding site becomes protected (3), and RING2, REP (4) and UPD are essentially unperturbed (Fig. 1b, Extended Data Fig. 3a).

Phosphorylation of parkin initiates release of REP and RING2 (4, 5), especially at later time points, but the phosphorylated Ubl also remains flexible and in exchange with solvent (1) (Fig. 1c, Extended Data Fig. 3b). The behaviour of phospho-Ubl changes markedly when a covalent, non-dischargeable E2-ubiquitin conjugate is added to the sample—now, the C-terminal RING2 peptide at the UPD interface is exposed to solvent (5), and the phosphorylated Ubl becomes protected (1), indicating the formation of a new interface (Fig. 1d, Extended Data Figs. 2, 3c). Finally, charging of the catalytic Cys of RING2 by ubiquitin was assessed using phosphorylated parkin covalently modified with the Ub-ABP ubiquitin-vinylsulfone (Ub-VS)^{5,13} (see Methods, Extended Data Fig. 2a-c). 'Charged' phospho-parkin reiterates the conformational changes observed in the phospho-parkin E2-Ub-bound sample (Fig. 1e, Extended Data Fig. 3d), showing that the ubiquitin-modified RING2 had been fully released from the parkin core (5). Overall, the HDX-MS experiments indicated that there were considerable rearrangements of Ubl and RING2, with loss of old and formation of new intramolecular interfaces on the parkin core (Fig. 1, Extended Data Figs. 2, 3).

Unexpectedly, a section of the linker between Ubl and UPD was protected during rebinding of the phospho-Ubl (6) (Fig. 1d, e, Extended Data Figs. 2, 3). This region of parkin, spanning amino acids 75–145,

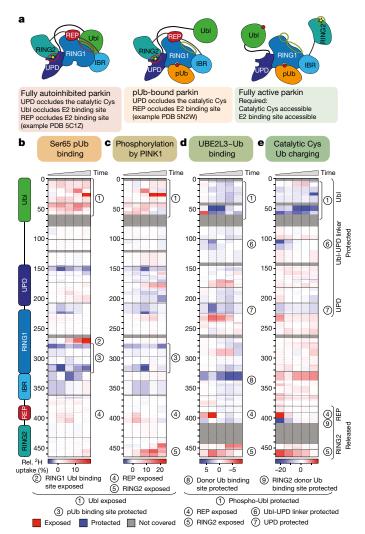


Fig. 1 | Domain rearrangements in parkin, resolved by HDX-MS. a, Cartoon of parkin activation. Left, parkin is autoinhibited by several mechanisms (red circles)⁷⁻⁹. Middle, binding of phospho-ubiquitin (pUb) to parkin releases the Ubl domain, but most mechanisms of autoinhibition remain^{5,6}. Right, after Ubl phosphorylation, parkin is fully active (green circles), but a structure of active parkin has not been reported. Also see Extended Data Fig. 1. **b**-**e**, HDX-MS difference maps with the shortest peptides covering any given region, coloured from blue (more protected from exchange compared to previous state) to red (more accessible to solvent exchange). Peptides for grey regions could not be analysed (see Extended Data Fig. 3). The five columns per sample indicate different time lengths for hydrogen-deuterium exchange (0.3 s, 3 s, 30 s, 300 s and 3,000 s). All experiments were performed with human full-length parkin, as technical triplicates. See Extended Data Figs. 2, 3 for raw data and structural mapping, respectively. b, Difference between parkin and parkin bound to phospho-ubiquitin. c, Difference between parkin-phosphoubiquitin and phospho-parkin-phospho-ubiquitin. d, Difference between phospho-parkin-phospho-ubiquitin, and phospho-parkin-phosphoubiquitin bound to a non-dischargeable UBE2L3-ubiquitin (Ub) complex (see Methods). e, Difference between phospho-parkin-phosphoubiquitin and phospho-parkin-phospho-ubiquitin charged with Ub-VS

has remained unstudied as it is disordered in full-length parkin⁷ and was removed in subsequent structures of human and rat parkin^{6,18,19}.

(see Methods).

The Ubl-UPD linker contains two connected, short sections of highly conserved residues that are flanked by a variable number of unconserved residues (Extended Data Fig. 4). A minimal linker is present in *Thamnophis sirtalis* (*Ts*) parkin (garter snake parkin, sequence identity to human parkin 73%, Extended Data Fig. 4), and *Ts*parkin

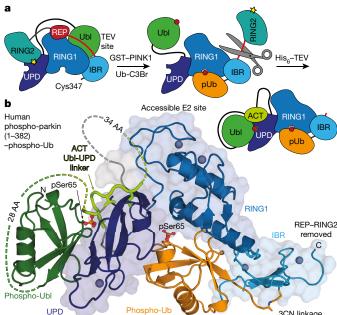


Fig. 2 | Structure of the phosphorylated parkin core. a, Schematic for obtaining a crystallizable phosphorylated parkin core. Scissors indicate the introduction of a TEV protease cleavage site after the IBR domain (amino acid 382). b, Crystal structure at 1.80 Å of the human phosphorylated parkin core lacking RING2, bound to phospho-ubiquitin. Phosphorylated residues are shown in ball-and-stick representation. A cartoon representation similar to a is shown to the right. Also see Extended Data Fig. 6 and Extended Data Table 1.

was used for comparative studies. HDX-MS revealed highly similar changes upon ubiquitin charging in phosphorylated *Ts*parkin when compared to human parkin (Extended Data Fig. 5a, with Fig. 1e). Moreover, limited proteolysis of full-length *Ts*parkin revealed that autoinhibited, unphosphorylated *Ts*parkin was cleaved first in the Ubl-UPD linker, whereas phosphorylated *Ts*parkin was cleaved first in the IBR-RING2 linker, and was not efficiently cleaved in the Ubl-UPD linker (Extended Data Fig. 5b). After cleavage of phospho-*Ts*parkin, RING2 was no longer stably associated with the parkin core (Extended Data Fig. 5c). Together, these data again strongly suggest that the unstudied Ubl-UPD linker becomes ordered in activated parkin, whereas REP and RING2 are dislodged, and RING2 becomes mobile.

We realized that crystallographic analysis of active parkin was likely to be impeded by a mobile RING2 domain, and this inspired new construct design. Parkin is insoluble when expressed without the RING2 domain (data not shown), probably owing to the exposed, hydrophobic UPD (Extended Data Fig. 1b). Hence, we engineered a tobacco etch virus (TEV) cleavage site into the IBR-RING2 linker (Fig. 2a, see Methods). This enabled us to remove the RING2 domain upon phospho-ubiquitin binding and Ubl phosphorylation (Extended Data Fig. 5d). Notably, Ub-VS-charged Tsparkin and Tsparkin lacking RING2 (Tsparkin∆RING2) displayed identical difference HDX-MS profiles, indicating that removal of the mobile RING2 had no effect on the remaining molecule (Extended Data Fig. 5e). For human parkin, the resulting covalent phospho-parkin∆RING2-phospho-ubiquitin (hereafter phospho-parkin-phospho-ubiquitin) complex was crystallized, and resulted in a 1.8 Å structure (Fig. 2b, Extended Data Table 1, Extended Data Fig. 6).

The structure of phospho-parkin–phospho-ubiquitin (Fig. 2b) revealed a near-identical organization of the parkin core (UPD–RING1–IBR) bound to phospho-ubiquitin, as compared to previous structures (r.m.s.d. 0.73 Å with human parkin–phospho-ubiquitin, PDB 5N2W⁶) (Extended Data Fig. 7a), and there were no large conformational changes in individual domains. Modelling of an open E2–Ub conjugate structure²⁸ reveals sensible interfaces (Extended Data Fig. 7b)

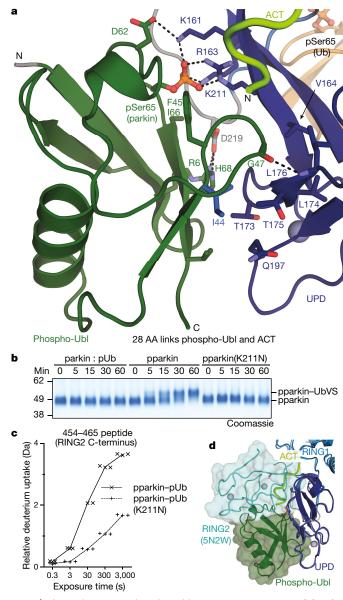


Fig. 3 | The parkin UPD-phospho-Ubl interaction. a, Structural detail of the binding site between parkin phospho-Ubl (green) and UPD (blue). Key residues are shown, and phospho-Ser65 is highlighted. Grey spheres indicate Zn atoms, and hydrogen bonds are shown as dotted lines. b, Ub-VS probe reactivity of the RING2 catalytic Cys residue with parkin-phospho-ubiquitin, phospho-parkin (pparkin), or phosphoparkin(K211N). The experiment was done in duplicate with identical results; for gel source data, see Supplementary Fig. 1. c, HDX-MS analysis of phospho-parkin-phospho-ubiquitin in comparison to phosphoparkin(K211N)-phospho-ubiquitin. The C-terminal peptide profiles are compared (see Extended Data Fig. 7d for overall data). The RING2 C terminus remains solvent-protected in the phospho-parkin(K211N) background. Technical triplicates are shown for all time points. d, Superposition of parkin-phospho-ubiquitin (PDB 5N2W⁶), and phospho-parkin-phospho-ubiquitin showing the relative positions of the RING2 (cyan surface) and phospho-Ubl (green surface), respectively, on the UPD domain.

and corroborates the ubiquitin binding site observed in HDX-MS⁶ (Fig. 1d, Extended Data Fig. 3c).

Notably, the phosphorylated Ubl domain was bound to the UPD, and had moved by more than 50 Å from its position in autoinhibited parkin (Figs. 2b, 3a, Extended Data Fig. 7a). The interface between phospho-Ubl and UPD is mediated by a common interaction site of ubiquitin-fold modifiers, the hydrophobic Ile44 patch of the Ubl, and engulfs the elongated UPD domain covering a surface of more than

800 Å²; this interface can be recapitulated in HDX-MS data (Fig. 1d, e, Extended Data Fig. 7c). Furthermore, the interaction places phosphorylated Ser65 into a positively charged pocket on the UPD (Fig. 3a). The phospho-acceptor pocket is lined by Lys161, Arg163 and Lys211, which contact the phosphate group and form four hydrogen bonds. We had previously noted this putative phosphate-acceptor binding site8, the importance of which is highlighted by two mutations found in patients with AR-JP (K211N and K161N)^{1,2} that also abrogate the function of parkin in mitophagy^{13,29}. Mechanistically, phosphorylated parkin with a K211N mutation blocking the phospho-acceptor pocket was no longer modified by Ub-VS⁵ (Fig. 3b). HDX-MS confirmed that phospho-parkin-phospho-ubiquitin(K211N) showed little sign of RING2 release and had the strongest relative solvent protection in the C terminus, where RING2 binds the UPD (Fig. 3c, Extended Data Fig. 7d). This indicated that the catalytic Cys of the RING2 domain remained inaccessible if phospho-Ubl was unable to interact with its UPD binding site, and explained how AR-JP-causing K211N or K161N mutations produce parkin variants that cannot be activated by Ser65 phosphorylation.

The position of the Ubl on the UPD overlaps only marginally with the position of RING2 in autoinhibited states of parkin, and while binding of both would lead to steric clashes (Fig. 3d, Extended Data Fig. 7a), the hydrophobic RING2 binding site (Extended Data Fig. 1b) would remain unusually exposed upon opening of the RING2-UPD interface. In our structure, clear electron density for a stretch of residues was apparent at the RING2-binding site of the UPD (Extended Data Fig. 6c), and we could unambiguously assign this density to the sequence corresponding to the first conserved region of the Ubl-UPD linker (Fig. 4, Extended Data Figs. 4, 6). In particular, residues Leu102, Val105 and Leu107 occupy pockets previously bound by RING2 residues Met458, Trp462 and Phe463 (Fig. 4a, b). Hence the Ubl-UPD linker shields the hydrophobic patch on the UPD that was opened by release of RING2. Indeed, similar to the K211N mutation, phosphoparkin with deletion of the first set of conserved linker residues $(\Delta 101-109)$ was unable to be charged by Ub-VS (Fig. 4c).

The linker provides additional contact points for the phospho-Ubl interface. Arg104 is located between two key hydrophobic residues, and contacts with its side chain the Ser65 loop in phospho-Ubl. Notably, parkin(R104W) is a mutation found in patients with AR-JP^{2,30}, and we would predict that this mutation would disrupt or misalign the observed hydrophobic interactions. A phospho-parkin(R104A) mutant was charged less efficiently by Ub-VS (Fig. 4d), showed slower E2–Ub discharge activity (Extended Data Fig. 8a, b) and reduced in vitro polyubiquitination activity (Fig. 4e), whereas its thermal stability remained unperturbed (Extended Data Fig. 8c).

Together, structural, biochemical and patient data confirm the crucial importance of the first conserved stretch of the Ubl–UPD linker for parkin activity, and define a new activating element, which we term ACT, in this understudied regulatory region of parkin, which also contains several phosphorylation sites (see further discussion in Extended Data Fig. 8d).

Our work resolves the activation mechanism of parkin, finally visualizing large-scale domain rearrangements and showing that the parkin Ubl switches between an inhibitory position in the unphosphorylated molecule to an activating position in phosphorylated parkin. Our data are consistent with a model in which the phosphorylated Ubl and the ACT element in the Ubl-UPD linker dislodge RING2 from its autoinhibited position, enabling it to be charged by E2–Ub, and ubiquitinate substrates in its vicinity independently of the parkin core (Fig. 4f). Notably, our model does not require parkin dimerization^{6,26}.

Our structure of an activated parkin core will inform drug discovery efforts that have set out to identify parkin activators. With the realization that the RING2–UPD interface opens and exposes a hydrophobic pocket, small molecules could be directed towards this interface. Such molecules may become particularly useful to restart mitophagy in patients with AR-JP who carry parkin variants that are not activated by PINK1-mediated Ubl phosphorylation.

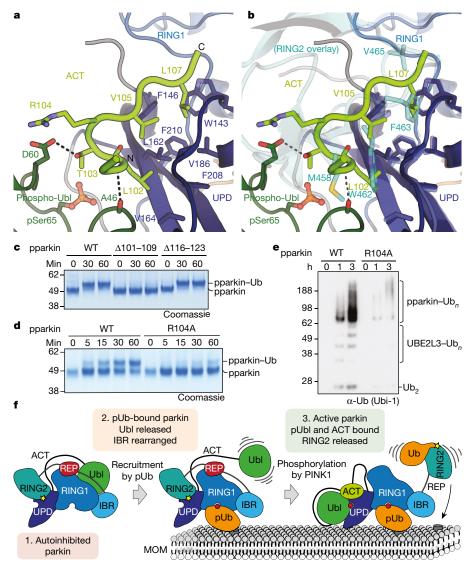


Fig. 4 | An activating element (ACT) in parkin. a, Structural detail of the ordered ACT within the parkin–phospho-Ubl–UPD linker. Three hydrophobic ACT residues bind the hydrophobic UPD groove, and polar ACT residues contact phospho-Ubl. b, Superposition of the ACT with RING2 (PDB 5N2W⁶, semi-transparent) in the same orientation as in a. Hydrophobic ACT residues mimic RING2 interactions. c, Ub-VS charging assay of phospho-parkin, and phospho-parkin variants lacking the ACT (Δ 101–109) or the second conserved hydrophobic linker sequence (Δ 116–123). Experiments were performed in duplicate with identical results; for gel source data, see Supplementary Fig. 1. d, Ub-VS charging assay as in c for wild-type (WT) phospho-parkin or the R104A mutant. Patients with parkin(R104W) suffer from AR-JP^{2,30}. Experiments were performed in duplicate with identical results; for gel source data, see Supplementary Fig. 1.

e, Activity of wild-type phospho-parkin and the R104A mutant in vitro, with UBE2L3 as the E2 enzyme. The reaction was resolved by SDS-PAGE and western blotted for ubiquitin. A representative gel of three independent experiments is shown. For source data, see Supplementary Fig. 1. f, Model of the sequential domain rearrangements required for full parkin activation, extended from a previous model³ (also see Fig. 1a). In autoinhibited parkin, the Ubl, REP and RING2 assume inhibitory positions. Binding of phospho-ubiquitin induces localization of parkin to the MOM, repositioning of the IBR domain and release of the Ubl domain. Phosphorylation of parkin allows the phospho-Ubl domain and ACT element to bind to the UPD, to replace and release the RING2 and REP, enabling ubiquitination of MOM substrates.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0224-x

Received: 4 May 2018; Accepted: 29 May 2018; Published online 6 June 2018.

- Corti, O., Lesage, S. & Brice, A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiol. Rev.* 91, 1161–1218 (2011).
- Pickrell, A. M. & Youle, R. J. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. Neuron 85, 257–273 (2015).
- 3. Harper, J. W., Ordureau, A. & Heo, J.-M. Building and decoding ubiquitin chains for mitophagy. *Nat. Rev. Mol. Cell Biol.* **19**, 93–108 (2018).

- Pickles, S., Vigié, P. & Youle, R. J. Mitophagy and quality control mechanisms in mitochondrial maintenance. *Curr. Biol.* 28, R170–R185 (2018).
- Wauer, T., Simicek, M., Schubert, A. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* 524, 370–374 (2015).
- Kumar, A. et al. Parkin-phosphoubiquitin complex reveals cryptic ubiquitinbinding site required for RBR ligase activity. *Nat. Struct. Mol. Biol.* 24, 475–483 (2017).
- Trempe, J.-F. et al. Structure of parkin reveals mechanisms for ubiquitin ligase activation. Science 340, 1451–1455 (2013).
- 8. Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112 (2013).
- Riley, B. E. et al. Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. Nat. Commun. 4, 1982 (2013).
- Koyano, F. et al. Ubiquitin is phosphorylated by PINK1 to activate parkin. Nature 510, 162–166 (2014).



- 11. Kane, L. A. et al. PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *J. Cell Biol.* **205**, 143–153 (2014).
- Kazlauskaite, A. et al. Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65. *Biochem. J.* 460, 127–139 (2014).
- Ordureau, A. et al. Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Mol. Cell* 56, 360–375 (2014).
- Wauer, T. et al. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. EMBO J. 34, 307–325 (2015).
- Kondapalli, C. et al. PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. Open Biol. 2, 120080 (2012).
- Iguchi, M. et al. Parkin-catalyzed ubiquitin-ester transfer is triggered by PINK1-dependent phosphorylation. J. Biol. Chem. 288, 22019–22032 (2013).
- 17. Shiba-Fukushima, K. et al. PINK1-mediated phosphorylation of the Parkin ubiquitin-like domain primes mitochondrial translocation of Parkin and regulates mitophagy. *Sci. Rep.* **2**, 1002 (2012).
- Sauvé, V. et al. A Ubl/ubiquitin switch in the activation of Parkin. EMBO J. 34, 2492–2505 (2015).
- Kumar, A. et al. Disruption of the autoinhibited state primes the E3 ligase parkin for activation and catalysis. EMBO J. 34, 2506–2521 (2015).
- Kazlauskaite, A. et al. Binding to serine 65-phosphorylated ubiquitin primes Parkin for optimal PINK1-dependent phosphorylation and activation. EMBO Rep. 16, 939–954 (2015).
- Ordureau, A. et al. Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. Proc. Natl Acad. Sci. USA 112, 6637–6642 (2015).
- Schubert, A. F. et al. Structure of PINK1 in complex with its substrate ubiquitin. Nature 552, 51–56 (2017).
- Ordureau, Á. et al. Dynamics of PARKIN-dependent mitochondrial ubiquitylation in induced neurons and model systems revealed by digital snapshot proteomics. *Mol. Cell* 70, 211–227.e8 (2018).
- Park, S., Foote, P. K., Krist, D. T., Rice, S. E. & Statsyuk, A. V. UbMES and UbFluor: novel probes for ring-between-ring (RBR) E3 ubiquitin ligase PARKIN. J. Biol. Chem. 292, 16539–16553 (2017).
- Pao, K.-C. et al. Probes of ubiquitin E3 ligases enable systematic dissection of parkin activation. Nat. Chem. Biol. 12, 324–331 (2016).
- Arkinson, C. & Walden, H. Parkin function in Parkinson's disease. Science 360, 267–268 (2018).

- Harrison, R. A. & Engen, J. R. Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. *Curr. Opin.* Struct. Biol. 41, 187–193 (2016).
- Lechtenberg, B. C. et al. Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. Nature 529, 546–550 (2016).
- Geisler, S. et al. PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. Nat. Cell Biol. 12, 119–131 (2010).
- Chaudhary, S. et al. Parkin mutations in familial and sporadic Parkinson's disease among Indians. Parkinsonism Relat. Disord. 12, 239–245 (2006).

Acknowledgements We thank beamline scientists at Diamond Light Source (DLS) for support at beamline I24; H. Ovaa and B.-T. Xin for H-Gly-VS hydrochloride; J. Pruneda, P. Elliott and M. Michel for help with crystallography and data collection; and members of the D.K. laboratory for reagents and discussions. Access to DLS was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570). The D.K. laboratory is supported by the Medical Research Council (U105192732), the European Research Council (724804), the Michael J. Fox Foundation and the Lister Institute for Preventive Medicine.

Reviewer information *Nature* thanks W. Harper, R. Youle and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions C.G. performed all experiments and analysed the data presented in this manuscript. S.L.M. and J.M.S. performed and analysed HDX-MS experiments. D.K. directed the project, analysed the data, acquired funding and wrote the manuscript with help from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0224-x.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.K. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Molecular biology. cDNA of *Thamnophis sirtalis* (*Ts*) parkin was obtained from GeneArt (Invitrogen) with codon-optimization for bacterial expression and cloned into a pOPIN-K vector³¹, using the In-Fusion HD Cloning Kit (Takara Clonetech). Human (*Hs*) parkin and *Pediculus humanus corporis* (*Ph*)PINK1 constructs were also expressed from a pOPIN-K vector, while UBE2L3 was expressed from a pGEX6 vector. *Hs*UBE1/PET21d was a gift from C. Wolberger (Addgene plasmid # 34965³²).

Site-directed mutagenesis was carried out using the QuikChange protocol with Phusion polymerase. A TEV cleavage site was introduced into the parkin constructs using the NEB Q5 Site-Directed Mutagenesis Kit (NEB). In *Ts*parkin, residues 368–374 (KSPGATA) were replaced by the ENLYFQS TEV cleavage sequence, while in *Hs*parkin residues 382–378 (EASGTTT) were replaced by the TEV cleavage sequence to yield cleavable constructs.

Protein purification. For parkin expression, *Escherichia coli* Rosetta2 pLacI cells (Novagen) were grown in $2\times TY$ medium at $37\,^{\circ}C$. At OD $_{600}=0.6$ the temperature was reduced to $18\,^{\circ}C$; expression was induced at OD $_{600}=0.8-1.0$ with $30\,\mu M$ IPTG and the medium supplemented with $200\,\mu M$ ZnCl $_2$. Cells were harvested after overnight growth at $18\,^{\circ}C$ and frozen at $-20\,^{\circ}C$.

For parkin purification, cells were resuspended in lysis buffer (300 mM NaCl, 10% (w/v) glycerol, 25 mM Tris (pH 8.5), 14.3 mM β -mercaptoethanol) supplemented with 2 mg/ml lysozyme, 0.2 mg/ml DNaseI and 80 µg/ml PMSF. The suspension was homogenized using an EmulsiFlex-C3 (Avestin) for two passes at $\sim\!15,000$ p.s.i. and cleared by centrifugation at 46,000g for 35 min at 4°C. The clarified lysate was applied to Amintra glutathione resin (Expedeon), resin washed with high salt buffer (25 mM Tris (pH 8.5), 500 mM NaCl, 10 mM DTT) and GST-fusion parkin cleaved from the resin overnight at 4°C with GST-3C protease.

Samples were eluted and resin washed with no-salt buffer (25 mM Tris (pH 8.5), 10 mM DTT). All following purification steps were carried out on an Äkta Pure system (GE Healthcare). Pooled fractions were subjected to anion-exchange chromatography on a 6-ml Resource Q column (GE Healthcare) with a 0–25% linear gradient from buffer A (25 mM Tris (pH 8.5), 10 mM DTT, 50 mM NaCl) to buffer B (25 mM Tris (pH 8.5), 10 mM DTT, 1,000 mM NaCl) over 15 column volumes. For phosphorylated parkin, the resulting sample was phosphorylated using a 1:100 molar ratio of GST–PhPINK1 in phosphorylation buffer (10 mM ATP, 10 mM MgCl₂, 200 mM NaCl, 50 mM Tris (pH 8.5), 10 mM DTT). PINK1 was subsequently removed by incubation with Amintra glutathione resin (Expedeon) and phosphorylated parkin purified using anion exchange chromatography as above. Finally, samples were subjected to size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) into buffer C (25 mM Tris (pH 8.5), 10 mM DTT, 200 mM NaCl).

In short, HsUBE1 was purified as follows. An N-terminal GST–Ub fusion protein was expressed and lysed in β -mercaptoethanol-free lysis buffer and applied to Amintra glutathione resin (Expedeon). Upon washing, the resin was equilibrated with 50 mM Tris (pH 8.5) and 2 mM ATP. HsUBE1 β -mercaptoethanol-free clarified lysate was generated as above, supplemented with 10 mM ATP and 10 MgCl₂ and incubated with the GST–Ub fusion-bound glutathione resin at room temperature for 30 min. The resin was then washed with DTT-free high salt buffer supplemented with 5 mM MgCl₂. HsUBE1 was eluted in DTT-containing buffer and protein-containing fractions were applied to anion-exchange and size-exclusion chromatography as above. UBE2L3, UBE2D3 and GST–PhPINK1 were purified as described previously³³.

Generation of non-dischargeable E2–Ub complex. UBE2L3 (C86K) and ubiquitin were stored in charging buffer (25mM CAPSO (pH 9.5), 20 mM MgCl₂, 150 mM NaCl). UBE2L3 (C86K) (450 μ M) was incubated with Ub (900 μ M) and HsUBE1 (2.5 μ M) in charging buffer supplemented with 10 mM ATP at 37 °C overnight. The resulting mixture was applied to size-exclusion chromatography as above in buffer C. Fractions containing UBE2L3–Ub were pooled, concentrated and again applied to size-exclusion chromatography to remove free UBE2L3.

Ub-VS generation and parkin coupling. Ub(1–75)–MesNa was prepared as described previously³⁴. H-Gly-VS hydrochloride was a kind gift from H. Ovaa and B.-T. Xin (Leiden University). Ub–MesNa, stored in buffer D (20 mM HEPES, 50 mM sodium acetate (pH 6.5), 75 mM NaCl) at ~20 mg/ml, was used to dissolve ~50 mg H-Gly-VS hydrochloride together with ~30 mg of N-hydroxysuccinamide (Fluka), acting as a catalyst. The pH was raised to 8.5 by addition of ~60 μ l of 4 M NaOH and reaction incubated at 37 °C. Reaction progress was monitored by LC–MS analysis. When the ratio of Ub(1–75)-VS to hydrolysed Ub–MesNa product was ~1:1, with a minimum formation of the doubly coupled, Ub(1–75)–VS–VS species, the reaction was quenched by addition of 20 μ l 12 M HCl (~30 min). The subsequent sample was diluted in 50 mM sodium acetate (pH 4.5) and applied to cation-exchange chromatography on a 1-ml MonoS column (GE Healthcare) with a 10–35% linear gradient between 50 mM sodium acetate (pH 4.5) containing 0 M and 1 M NaCl, respectively. The resulting fractions were analysed by LS–MS

and Ub(1-75)–VS containing fractions were pooled and applied to size-exclusion chromatography as above in buffer D.

For quantitative parkin–Ub-VS coupling, phospho-parkin was purified as above where 10 mM DTT in buffer C was replaced with 5 mM TCEP. Phospho-parkin and Ub-VS were mixed at a 1:3 molar ratio and incubated at room temperature. Reaction progress was monitored by LC–MS analysis and, upon completion, the reaction was quenched by addition of DTT ($\sim\!30$ min). The resulting sample was purified using size-exclusion chromatography (buffer C).

Mass-spectrometry analysis. LC–MS analysis was carried out on an Agilent 1200 Series chromatography system coupled to an Agilent 6130 Quadrupole mass spectrometer. Samples were eluted from a Phenomenex Jupiter column (5 ml, 300 Å, C4 column, 150 \times 2.0 mm) using an acetonitrile gradient + 0.2% (v/v) formic acid. Protein was ionized using an ESI source (3 kV ionization voltage), and spectra were analysed in positive ion mode with a mass range between 400 and 2,000 m/z. Averaged spectra were deconvoluted using Promass (Novatia, LLC) and plotted using GraphPad Prism (version 7).

Limited proteolysis. *Ts*parkin, phospho-*Ts*parkin, phospho-*Ts*parkin-phospho-ubiquitin, and phospho-*Ts*parkin-phospho-ubiquitin charged with Ub-VS were purified as described above. A 1 mg/ml protein solution was mixed with 5 μ g/ml solution of elastase from the Proti-Ace Kit (Hampton research) and incubated for 1 h at room temperature. The reactions were quenched by addition of DTT- and iodoacetamide-containing LDS buffer and resolved on a 4–12% SDS NuPAGE gradient gels (Invitrogen) and stained with Instant Blue SafeStain (Expedeon).

Hydrogen–deuterium exchange mass spectrometry (HDX-MS). Complexes were formed on ice and incubated for 30 min to give a final parkin concentration of 10 μ M. Deuterium-exchange reactions of parkin and the different complexes were initiated by diluting the protein in D2O (99.8% (v/v) D2O ACROS, Sigma) in 25 mM Tris (pH 8.5), 200 mM NaCl, 1 mM TCEP to give a final D2O percentage of $\sim\!95\%$. For all experiments, deuterium labelling was carried out at 23 °C (unless otherwise stated) at five time points: 0.3 s (3 s on ice), 3 s, 30 s, 300 s and 3,000 s, in technical triplicate. The labelling reaction was quenched by the addition of chilled 2.4% (v/v) formic acid in 2 M guanidinium hydrochloride and immediately frozen in liquid nitrogen. Samples were stored at -80 °C before analysis.

The quenched protein samples were rapidly thawed and subjected to proteolytic cleavage with pepsin followed by reversed phase HPLC separation. In brief, the protein was passed through an Enzymate BEH immobilized pepsin column, 2.1×30 mm, $5~\mu m$ (Waters, UK) at $200~\mu l/min$ for 2 min, the peptic peptides were trapped and desalted on a 2.1×5 mm C18 trap column (Acquity BEH C18 Vanguard pre-column, $1.7~\mu m$, Waters). Trapped peptides were subsequently eluted over 11 min using a 3-43% gradient of acetonitrile in 0.1% (v/v) formic acid at $40~\mu l/min$. Peptides were separated on a reverse phase column (Acquity UPLC BEH C18 column $1.7~\mu m$, $100~mm\times1~mm$; Waters) and detected on a SYNAPT G2-Si HDMS mass spectrometer (Waters) over an m/z of 300 to 2,000, with the standard electrospray ionization (ESI) source with lock mass calibration using [Glu1]-fibrino peptide B (50 fmol/ μl). The mass spectrometer was operated at a source temperature of $80~\rm ^{\circ}C$ and a spray voltage of $2.6~\rm kV$. Spectra were collected in positive ion mode.

Peptide identification was performed by MS^{e35} using an identical gradient of increasing acetonitrile in 0.1% (v/v) formic acid over 11 min. The resulting MS^e data were analysed using Protein Lynx Global Server software (Waters, UK) with an MS tolerance of 5 ppm.

Mass analysis of the peptide centroids was performed using DynamX software (Waters). Only peptides with a score >6.4 were considered. The first round of analysis and identification was performed automatically by the DynamX software, however, all peptides (deuterated and non-deuterated) were manually verified at every time point for the correct charge state, presence of overlapping peptides, and correct retention time. Deuterium incorporation was not corrected for back-exchange and represents relative, rather than absolute changes in deuterium levels. Changes in H/D amide exchange in any peptide may be due to a single amide or a number of amides within that peptide.

Protein preparation for crystallization. TEV-cleavable parkin was purified as described above. An anion-exchange purified parkin sample (step 1) was incubated with Ub-C3Br in a 1:4 molar ratio and GST-PhPINK1 in a 9:1 molar ratio in phosphorylation buffer (10 mM ATP, 10 mM MgCl₂, 200 mM NaCl, 50 mM Tris (pH 8.5), 10 mM DTT) at a final parkin concentration of ~70 µM, yielding a phosphorylated, phospho-ubiquitin conjugated, TEV-cleavable parkin sample (step 2). GST-PhPINK1 was subsequently removed using Amintra glutathione resin (Expedeon). The sample was subjected to His₆-TEV cleavage overnight at 4 °C (step 3). His₆-TEV was removed using Ni-NTA agarose (Qiagen), sample diluted in buffer A and applied to anion-exchange and size-exclusion chromatography as described above.

To generate Ub-C3Br, Ub(1–75)–MesNa was prepared as described previously 34 . Ub–MesNa, stored in buffer D was incubated with 0.2 g/ml 3-bromopropylamine hydrobromide (Fluka) dissolved in PBS (pH 4.8) at 2:1 molar ratio



with final Ub(1–75)–MesNa concentration of 445 μM . The coupling was carried out on ice for 30 min following addition of 50 μl 4 M NaOH to raise the pH to 10.5. The reaction was quenched by addition of 12 μl of 12 M HCl and sample buffer exchanged using a disposable PD-10 desalting column (GE Healthcare) into buffer C.

Crystallization. Initial crystals were found from crystallization experiments carried out at 18 °C in a 96-well sitting drop vapour diffusion plates in the MRC format (Molecular Dimensions) by mixing 100 nl of 4 mg/ml protein solution with 100 nl reservoir solution. The crystallization condition of 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, 12.5% (v/v) MPD, 0.03 M of each sodium nitrate, disodium hydrogen phosphate, ammonium sulfate, 0.1 M MOPS/ HEPES-Na (pH 7.5) was found from the MORPHEUS screen (Molecular Dimensions). Seeds were obtained from a fine screen and streak seeding was carried out in a hanging drop format from an 8 mg/ml protein solution. Larger crystals were obtained after 6 days in the original crystallization condition. Crystals were soaked in mother-liquor supplemented with 10% (v/v) glycerol before vitrification in liquid nitrogen.

Data collection, phasing and refinement. Diffraction data were collected at the Diamond Light Source, beamline I-24 (0.9686 Å, 100 K), and processed using DIALS³⁶. The crystal structure was determined by molecular replacement in Phaser³⁷ using the structure of the human parkin core (PDB 5N2W⁶) truncated after the IBR, as well as a human parkin Ubl structure (PDB 5C1Z¹⁹). The structure was built at 1.80 Å, in multiple rounds of model building in Coot³⁸ and refinement in PHENIX³⁹. Phenix ReadySet-derived geometry restraints for the 3CN warhead were used, with external restraints defining the linkage points. Final Ramachandran statistics: 98.9% favoured, 1.1% allowed, and 0% outliers. Structural figures were generated using PyMol (http://www.pymol.org). Data collection and refinement statistics can be found in Extended Data Table 1.

Parkin activity assays. *Ub-VS conjugation assays.* Indicated parkin variants stored in either DTT- or TCEP-containing buffer were incubated with Ub-VS that was prepared as described above. The reactions were quenched at indicated time points by addition of DTT- and iodoacetamide-containing LDS buffer and resolved on a 4–12% SDS NuPAGE gradient gels (Invitrogen) and stained with Instant Blue SafeStain (Expedeon).

Parkin assembly assays. Wild-type or R104A phospho-parkin (5 μM) were incubated in ubiquitination buffer (30 mM HEPES (pH 7.5), 100 mM NaCl, 10 mM ATP, 10 mM MgCl $_2$) with HsUBE1 (0.2 μM), UBE2L3 (2 μM) and Ub (20 μM). The reactions were quenched at the indicated time points by addition of DTT- and iodoacetamide-containing LDS buffer and resolved on a 4–12% SDS NuPAGE gradient gels (Invitrogen) and transferred to a PVDF membrane (BioRad). Membranes were blocked in a 5% (w/v) milk solution in PBS-T (PBS + 0.1% (v/v) Tween-20) for 30 min and incubated overnight at 4 °C with a ubiquitin-recognizing antibody (Ubi-1, NB300-130, Novus Biologicals) in 5% (w/v) BSA in PBS-T and 0.1% (w/v) sodium azide. The membrane was then washed with PBS-T, incubated for 1 h at room temperature with anti-mouse [gG-HRP (NXA931, GE Healthcare) in 5% (w/v) milk in PBS-T, washed in PBS-T and visualized using the Amersham Western Blotting Detection Reagent (GE Healthcare) and a ChemiDoc Touch Imaging System (BioRad).

E2–Ub discharge assays. The UBE2D3–Ub conjugate was generated by incubating UBE2D3 (20 $\mu M)$ with HsUBE1 (20 nM) and Ub (80 $\mu M)$ in ubiquitination buffer supplemented with 5 μM CaCl $_2$ at 37 °C for 10 min. To remove remaining

ATP, 0.5 U of Apyrase (NEB) was added and the reaction incubated at 30 $^{\circ}\mathrm{C}$ for 30 min.

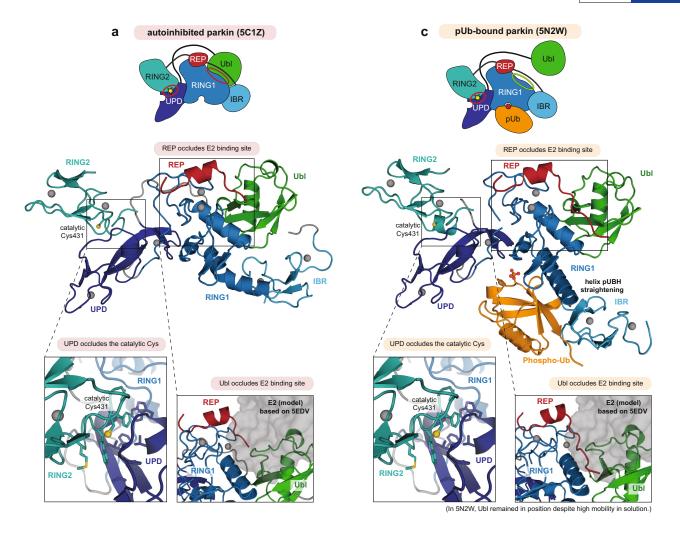
The discharge reaction was studied by addition of 1 μM wild-type or R104A phospho-parkin to a diluted charging reaction mixture (final UBE2D3 concentration was 9 μM). The reactions were quenched at indicated time points by addition of DTT-free LDS buffer, while a final sample was collected at 11 min in DTT-containing LDS buffer to assess the extent of isopeptide-linked UBE2D3–Ub species formation. Samples were resolved on 4–12% SDS NuPAGE gradient gels (Invitrogen) and stained with Instant Blue SafeStain (Expedeon). The gel band intensity was quantified in ImageJ by isolating the specific intensity of the UBE2D3~Ub thioester band as indicated, subtracting the background of the final reduced sample and normalized within each reaction.

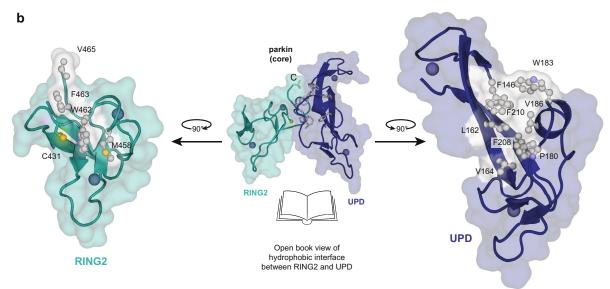
Thermal denaturation assays. Protein melting curves were recorded on a Corbett RG-6000 real time PCR cycler (30 °C to 85 °C with 7 s per 0.5 °C). Samples contained 4 μ M parkin protein and 4×SYPRO orange in ubiquitination buffer + 5 mM TCEP. Melting curves were obtained as the maxima of d*F*/d*T* versus *T* plots. All data were recorded in triplicate.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Coordinates and structure factors have been deposited with the Protein Data Bank under accession code 6GLC. Uncropped versions of all gels are displayed in Supplementary Fig. 1. All reagents and data are available upon reasonable request from the corresponding author.

- Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* 35, e45 (2007)
- Berndsen, C. E. & Wolberger, C. A spectrophotometric assay for conjugation of ubiquitin and ubiquitin-like proteins. *Anal. Biochem.* 418, 102–110 (2011).
- Gladkova, C. et al. An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. EMBO J. 36, 3555–3572 (2017).
- Wilkinson, K. D., Gan-Erdene, T. & Kolli, N. Derivitization of the C-terminus of ubiquitin and ubiquitin-like proteins using intein chemistry: methods and uses. Methods Enzymol. 399, 37–51 (2005).
- Silva, J. C. et al. Quantitative proteomic analysis by accurate mass retention time pairs. Anal. Chem. 77, 2187–2200 (2005).
- Waterman, D. G. et al. Diffraction-geometry refinement in the DIALS framework. Acta Crystallogr. D 72, 558–575 (2016).
- McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).
- Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94–106 (2011).
- Yamamoto, A. et al. Parkin phosphorylation and modulation of its E3 ubiquitin ligase activity. J. Biol. Chem. 280, 3390–3399 (2005).
- Lu, X. et al. Mitophagy controls beige adipocyte maintenance through a Parkin-dependent and UCP1-independent mechanism. Sci. Signal. 11, eaap8526 (2018).
- Avraham, E., Rott, R., Liani, E., Szargel, R. & Engelender, S. Phosphorylation of Parkin by the cyclin-dependent kinase 5 at the linker region modulates its ubiquitin-ligase activity and aggregation. J. Biol. Chem. 282, 12842–12850 (2007)
- Durcan, T. M. et al. USP8 regulates mitophagy by removing K6-linked ubiquitin conjugates from parkin. EMBO J. 33, 2473–2491 (2014).

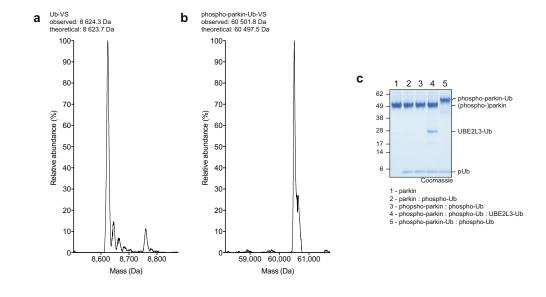


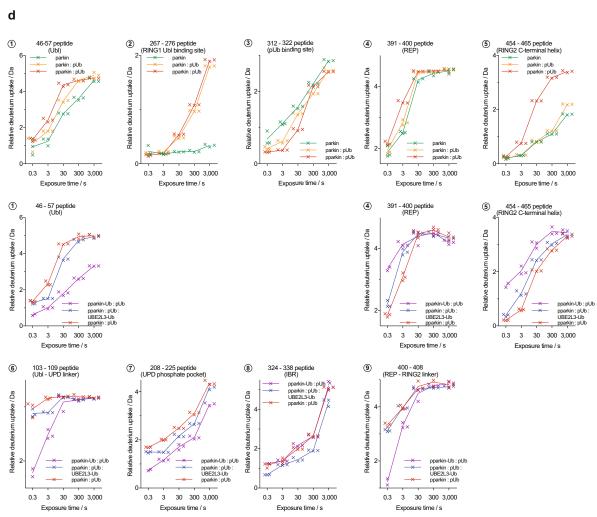


Extended Data Fig. 1 | Mechanisms of parkin autoinhibition.

a, Structure of autoinhibited, full-length human parkin (PDB 5C1Z¹⁹) shown schematically (top, as in Fig. 1a) and in cartoon representation in the same colours. Two insets show the UPD–RING2 interface (with Cys431 shown in ball-and-stick representation), and the blocked E2 binding site (with the E2 position, modelled according to PDB 5EDV²⁸, shown as grey surface). Zn ions are shown as grey spheres. **b**, An 'openbook' view of the UPD–RING2 interface, with hydrophobic residues

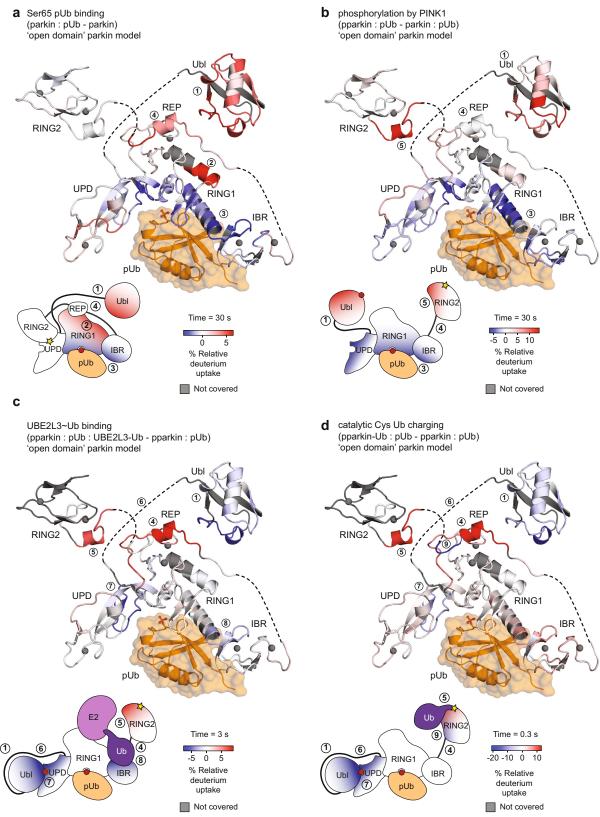
coloured white on each surface. c, Structure of phospho-ubiquitin bound to full-length parkin (PDB 5N2W⁶) as in a. Phospho-ubiquitin binding leads to helix straightening, and IBR domain repositioning, which releases the Ubl domain for phosphorylation^{5,6}. In the shown structures of unphosphorylated parkin, the Ubl and REP (red) inhibit E2 binding, and the RING2–UPD interface is intact, with Cys431 being inaccessible. The Ubl–UPD linker was removed from crystallized constructs in a and c^{6,19}.





Extended Data Fig. 2 | Sample preparation for HDX-MS and selected raw data. a, Representative LC-MS spectrum of the prepared Ub-VS probe (see Methods). Experiment was performed in duplicate. b, Representative LC-MS spectrum of Ub-VS-reacted phospho-parkin. Experiment was performed in duplicate. c, Samples used in HDX-MS analysis. In HDX-MS, non-covalent complexes with phospho-ubiquitin were used. Covalent complexes are indicated with a dash and non-covalent

complexes by a colon. This is representative of at least three independent experiments; for gel source data, see Supplementary Fig. 1. \mathbf{d} , Relative deuterium uptake (in Da) is shown for exemplary selected peptides across the parkin molecule, over the timecourse of the experiment. Each point for the technical replicate experiments is shown. Data points were taken at identical time points, but are offset on the x axis for clarity.



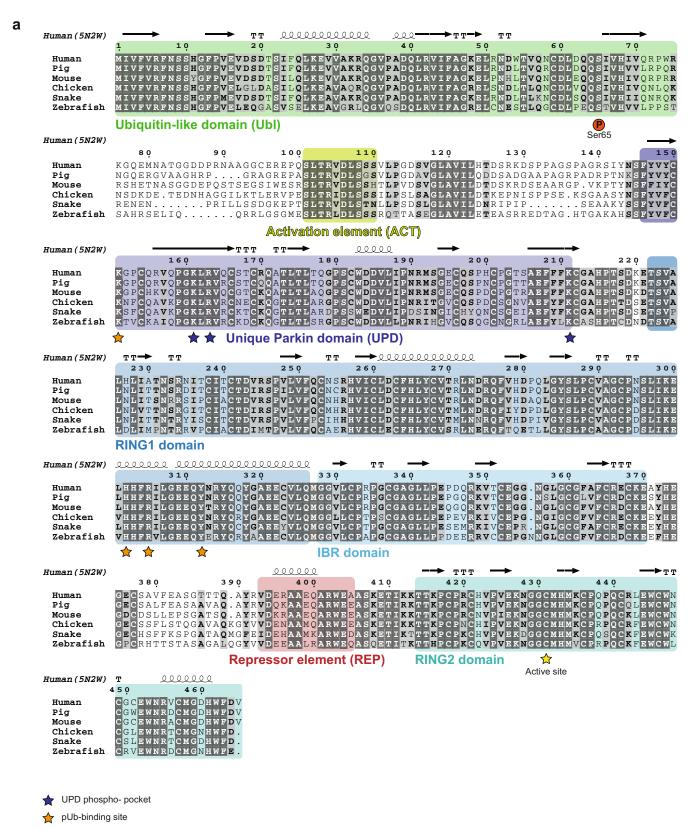
Extended Data Fig. 3 | See next page for caption.



Extended Data Fig. 3 | Graphical representation of HDX-MS data.

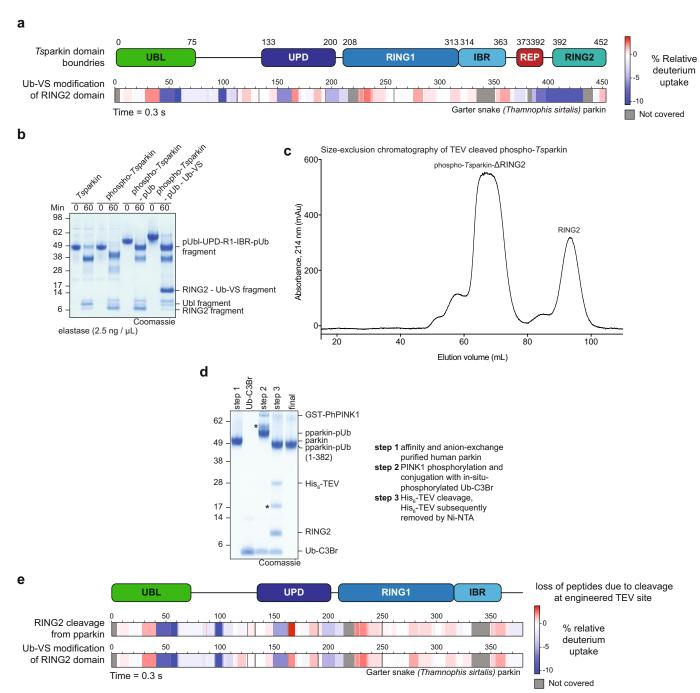
Data from HDX-MS experiments (Fig. 1b–e) were plotted onto a stylized 'open domain' model of parkin, with identical colouring (blue, more protected from solvent exchange compared to previous state; red, less protected from solvent exchange compared to previous state). Grey regions correspond to peptides that were not covered or could not be analysed owing to modification. Schematic domain representations indicate an average change of the corresponding interfaces across all time points. White regions indicate no change. a, Parkin compared to parkin–phospho-ubiquitin. b, Parkin–phospho-ubiquitin compared to phospho-parkin–phospho-ubiquitin compared to phospho-parkin–phospho-ubiquitin in complex with an isopeptide UBE2L3–Ub thioester mimetic (see Methods). This experiment confirmed a previously reported binding site for the E2-conjugated ubiquitin on the

RBR^{6,28} (8). **d**, Phospho-parkin–phospho-ubiquitin compared to Ub-VS-reacted phospho-parkin–phospho-ubiquitin. Reaction with Ub-VS leads to modification of the catalytic Cys431-containing-peptide, generating non-identical peptides precluding comparison by HDX-MS. Low coverage of the RING2 domain can be explained by ubiquitin resistance to pepsin cleavage, leading to protection of the linked RING2 domain and subsequent peptide loss. To allow comparison, these peptides were also omitted from analysis of the UBE2L3–Ub-bound sample. In **c** and **d**, the structure representation is deceiving because REP and RING2 are highly mobile and are no longer bound to the parkin core. Indeed, the high hydrogen–deuterium exchange in the REP sequence in active parkin (Fig. 1d, e; peptide (4) in Extended Data Fig. 2d) indicates an additional loss of secondary structure in this helical element when REP and RING2 are released.



Extended Data Fig. 4 | A conserved linker between Ubl and UPD. Sequence alignment of parkin, with domains coloured corresponding to $5N2W^6$ as in Extended Data Fig. 1. Phosphate binding pockets are labelled. The linker region between Ubl and UPD (amino acids 76–143) contains two strings of highly conserved residues. Residues upstream and

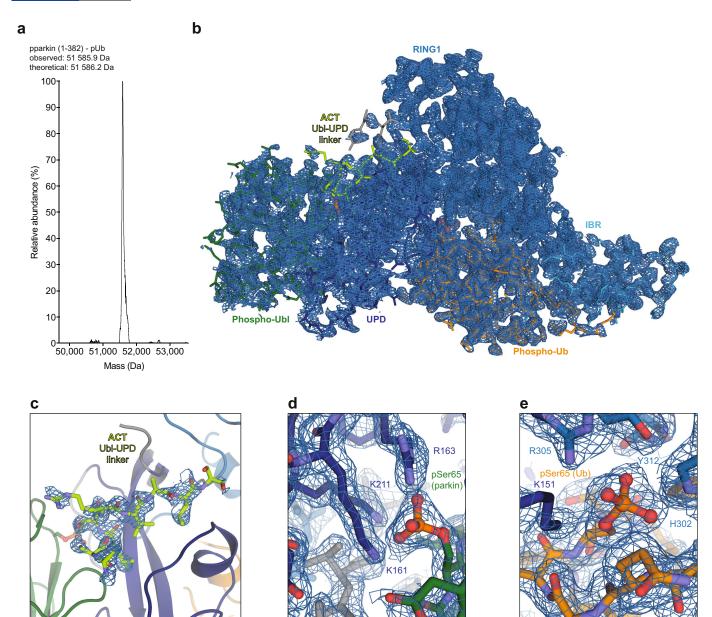
downstream of the conserved region are unconserved both in sequence and linker length. *Ts*parkin shows the smallest number of residues in the linker (upstream, 25 amino acids in human parkin, 18 amino acids in *Ts*parkin; downstream, 18 amino acids in human parkin, 11 amino acids in *Ts*parkin). See also Extended Data Fig. 8d.



Extended Data Fig. 5 | See next page for caption.

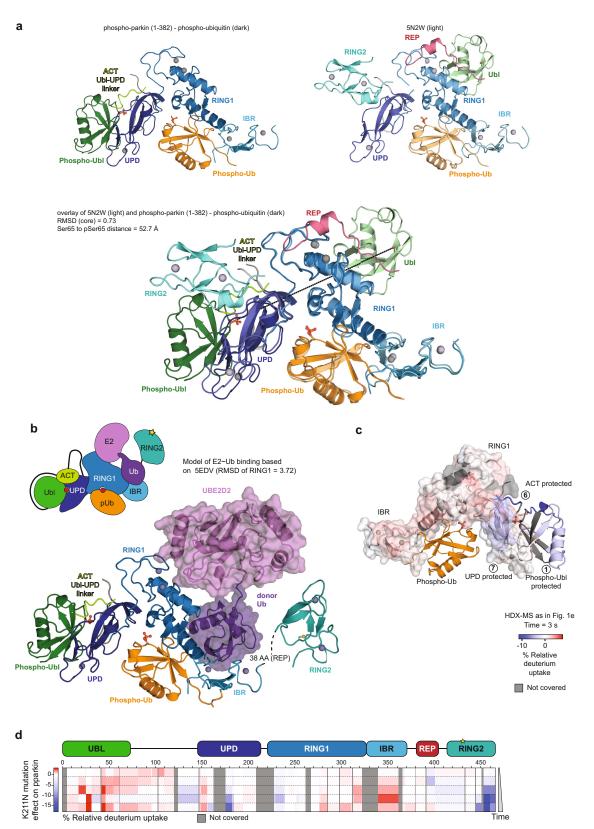
Extended Data Fig. 5 | Tsparkin and pre-crystallization biochemistry for human parkin. a, HDX-MS experiment comparing phospho-Tsparkin reacted with phospho-ubiqutin-C3Br and phospho-Tsparkin reacted with phospho-ubiquitin-C3Br and Ub-VS with identical colouring (blue, more protected from solvent exchange; red, less protected from solvent exchange; grey, not covered in all of the compared states, see Fig. 1). The experiment was performed in technical triplicate. The Tsparkin profile is highly similar to the profile of human parkin in an analogous state (Fig. 1e). Higher peptide resolution in this sample reveals protection of the RING2 interface by reacted Ub-VS, but the C terminus of RING2 that binds to the UPD interface is surface exposed. Both phospho-Ubl and the Ubl-UPD linker are protected in activated parkin. b, Limited proteolysis of Tsparkin with elastase, in different stages of activation. In unphosphorylated, autoinhibited Tsparkin, the Ubl is cleaved off in the Ubl-UPD linker. In activated forms of Tsparkin (phospho-Tsparkin, phospho-Tsparkin reacted with phospho-ubiquitin-C3Br, phospho-Tsparkin reacted with phospho-ubiquitin-C3Br and Ub-VS), the RING2 is readily cleaved off, while the Ubl is not efficiently removed. This suggests that the Ubl-UPD linker is not accessible in activated forms of Tsparkin. A representative gel from three independent experiments is shown. For gel source data, see Supplementary Fig. 1. c, A TEV cleavage

site was introduced after the IBR domain, so that after activation by phospho-ubiquitin and Ubl-phosphorylation, the released RING2 domain can be removed. Once removed, RING2 is no longer stably associated with the remaining parkin core. Shown is a gel filtration profile illustrating this point. A representative profile from three independent experiments is shown. d, SDS-PAGE analysis of sample preparation process (see Methods). Asterisk denotes ubiquitin probe (Ub-C3Br)reacted material that modifies the RING2 catalytic Cys, which explains the cleaved, probe-reacted RING2 band (asterisk in step 3). A representative gel from three independent experiments is shown. For gel source data, see Supplementary Fig. 1. e, HDX-MS experiment on Tsparkin, comparing phospho-Tsparkin reacted with phospho-ubiquitin-C3Br with phospho-Tsparkin reacted with phospho-ubiquitin-C3Br and Ub-VS (bottom) or with RING2-TEV-cleaved phospho-Tsparkin reacted with phosphoubiquitin-C3Br (top), coloured as in a. Identical profiles were obtained, showing that RING2 removal has no effect on the activated core of parkin. This further indicates that RING2 acts independently of the parkin core upon full activation. Notably, in both comparisons, we observed concomitant protection of phospho-Ubl and the Ubl-UPD linker. The experiment was performed in technical triplicate.



Extended Data Fig. 6 | Quality control and electron density maps for human phospho-parkin-phospho-ubiquitin. a, LC-MS spectrum of crystallized human phospho-parkin (amino acids 1–382) bound to phospho-ubiquitin. This is representative of two independent experiments. b, Composite omit map (generated with simulated annealing) shown for the single complex in the asymmetric unit. $2|F_0|-|F_c|$ electron density is shown at 1σ . c, Electron density as in b for

the Ubl–UPD linker. **d**, Electron density as in **b** for the Ser65 phospho-Ubl binding site on the UPD linker. **e**, Electron density as in **b** for the Ser65 phospho-Ub binding site. As we are missing electron density for disordered regions in the Ubl–ACT and ACT–UPD linkers, we cannot exclude the possibility that phospho-Ubl may interact *in trans* with a neighbouring parkin molecule. Also see Extended Data Table 1.



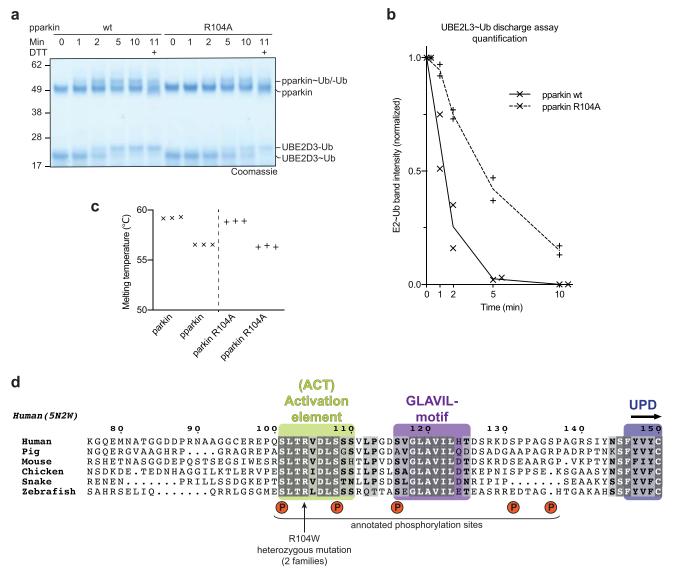
Extended Data Fig. 7 | See next page for caption.



Extended Data Fig. 7 | The phospho-Ubl binding site on the UPD.

a, Side-by-side view of phospho-parkin-phospho-ubiquitin (left) and parkin-phospho-ubiquitin (PDB 5N2W⁶, right), and superposition of both (below). The green Ubl domain changes position by >50 Å.
b, E2-Ub from the structure of the HOIP RBR domain in complex with UBE2D2-Ub²⁸ was modelled onto phospho-parkin-phospho-ubiquitin, by superposition of the RING1 domains of each complex. The E2-conjugated ubiquitin molecule in the 'open' conformation binds to the previously recognized cryptic ubiquitin binding interface on RING1-IBR⁶. The contact points correlate with HDX-MS data (Fig. 1d, Extended

Data Figs. 2, 3c). c, HDX-MS data from Fig. 1e were plotted onto the phospho-parkin-phospho-ubiquitin structure with identical colouring (blue, more protected from solvent exchange; red, less protected from solvent exchange; grey, not covered in all of the compared states, compare with Fig. 1). Protected regions on UPD match the observed phospho-Ubl interface. d, HDX-MS experiments comparing parkin with a mutation in the phospho-acceptor binding site on the UPD (phospho-parkin(K211N)-phospho-ubiquitin) compared with phospho-parkin-phospho-ubiquitin, coloured as in c. The mutant is unable to protect the Ubl, and to release RING2 and REP. Experiments were done as technical triplicate.



Extended Data Fig. 8 | A regulatory role of the parkin Ubl-UPD linker. a, b, E2 discharge assay resolved on a Coomassie stained SDS-PAGE gel (a) and quantified from band intensities (b) for phospho-parkin and phospho-parkin(R104A). This is representative of at least two independent experiments; for gel source data, see Supplementary Fig. 1. The mutation in the ACT element leads to a reduction in discharge activity, suggesting that the residue is required to dislodge RING2 from the parkin core. c, Parkin(R104A) is equally stable as wild-type parkin, in the unphosphorylated or phosphorylated form. Thermal denaturation experiments were performed as technical triplicate. d, Sequence detail of the Ubl-UPD linker, which contains the ACT element described here. In the ACT element as bound to phospho-parkin-phosphoubiquitin, the positions of two annotated (in PhosphoSitePlus) parkin phosphorylation sites, Ser101 and Ser108, are resolved. Phosphorylation of Ser101 decreases parkin activity⁴⁰, which is probably explained by phosphorylation preventing phospho-Ubl and/or linker binding to the UPD. It is hence highly likely that phosphorylation of parkin on these residues provides additional layers of parkin regulation that remain to be uncovered in future work. As an example, parkin phosphorylation by PKA was recently reported to be a mechanism of parkin inhibition in beigeto-white adipocyte transition, although phosphorylation sites remained unclear⁴¹. Residues before the ACT element (amino acids 73-99) and after the ACT element (amino acids 109-142) are disordered in our structure. The last ordered residue, Ser108, is tantalizingly close to the REP binding site as well as to the phospho-ubiquitin binding pocket, but disorder suggests that clear binding sites for other conserved linker residues, in particular for the parkin GLAVIL motif, are not present. HDX-MS also does not reveal additional protection of the linker, even when the E2-Ub conjugate is bound, suggesting that the GLAVIL motif may not bind the E2 (Fig. 1d, Extended Data Figs. 2, 3c). On the other hand, there are at least three additional annotated phosphorylation sites, Ser116, Ser131 and $Ser136^{15,40,42,43}$, suggesting that the second part of the linker may also be regulated. Phosphorylation on these residues could change the ability of the disordered parts of the linker to interact with parkin *in cis*. For example, we would speculate that a phosphorylated Ser116 could for example, reach the phosphate binding pocket occupied by phospho-Ser65 of ubiquitin. Alternatively, the remaining Ubl-UPD linker may be important for substrate recruitment, or involved in other, PINK1independent mechanisms of parkin activation.



Extended Data Table 1 \mid Data collection and refinement statistics

	phospho-parkin (1-382) - phospho-Ub
Data collection	
Space group	P 3 ₂ 2 1
Cell dimensions	
a, b, c (Å)	83.93, 83.93, 105.12
α , β , γ (°)	90, 90, 120
Resolution (Å)	72.69 –1.80 (1.84 – 1.80)
R _{merge}	0.065 (0.773)
1/ ol	13.80 (2.40)
Completeness (%)	100.00 (99.30)
Redundancy	6.7 (6.7)
Refinement	
Resolution (Å)	59.79 –1.80
No. reflections / test set	40229 / 2020
R _{work} / R _{free}	0.180 / 0.205
No. atoms	
Protein	3039 (398 aa)
Ligand/ion	41
Water	165
B-factors	
Protein	45.05
Ligand/ion	66.19
Water	46.94
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.19

^{*}Values in parentheses are for highest-resolution shell.



Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action

Yan Yan^{1,8}, Qikun Liu^{2,8}, Xin Zang^{3,4,8}, Shuguang Yuan⁵, Undramaa Bat-Erdene¹, Calvin Nguyen², Jianhua Gan⁶, Jiahai Zhou^{3,4*}, Steven E. Jacobsen^{2*} & Yi Tang^{1,7*}

Bioactive natural products have evolved to inhibit specific cellular targets and have served as lead molecules for health and agricultural applications for the past century¹⁻³. The post-genomics era has brought a renaissance in the discovery of natural products using synthetic-biology tools⁴⁻⁶. However, compared to traditional bioactivity-guided approaches, genome mining of natural products with specific and potent biological activities remains challenging⁴. Here we present the discovery and validation of a potent herbicide that targets a critical metabolic enzyme that is required for plant survival. Our approach is based on the co-clustering of a selfresistance gene in the natural-product biosynthesis gene cluster⁷⁻⁹ which provides insight into the potential biological activity of the encoded compound. We targeted dihydroxy-acid dehydratase in the branched-chain amino acid biosynthetic pathway in plants; the last step in this pathway is often targeted for herbicide development¹⁰. We show that the fungal sesquiterpenoid aspterric acid, which was discovered using the method described above, is a sub-micromolar inhibitor of dihydroxy-acid dehydratase that is effective as a herbicide in spray applications. The self-resistance gene astD was validated to be insensitive to aspterric acid and was deployed as a transgene in the establishment of plants that are resistant to aspterric acid. This herbicide-resistance gene combination complements the urgent ongoing efforts to overcome weed resistance¹¹. Our discovery demonstrates the potential of using a resistance-gene-directed approach in the discovery of bioactive natural products.

Weeds are a major source of crop losses, and the evolution of herbicide resistance in weeds has led to an urgent need for new herbicides with novel modes of action 11-14. The branched-chain amino acid (BCAA) biosynthetic pathway is essential for plant growth¹⁰. It is not present in animals and is therefore a validated target for highly specific weed-control agents¹⁰. The BCAA biosynthetic pathway in plants is carried out by three enzymes: acetolactate synthase (ALS), acetohydroxy acid isomeroreductase (KARI), and dihydroxyacid dehydratase (DHAD) (Fig. 1a). Given the success of targeting ALS for herbicide development¹¹, it is notable that no herbicide that targets either of the other two enzymes has been developed. DHAD is an essential and highly conserved enzyme among plant species that catalyses β -dehydration reactions to yield α -keto acid precursors to isoleucine, valine and leucine^{15,16} (Extended Data Fig. 1a, Supplementary Fig. 1). Efforts towards synthetic DHAD inhibitors resulted in compounds with submicromolar inhibition constants (K_i) , however, the compounds have no reported in planta activity¹⁷ (Extended Data Fig. 1b).

Filamentous fungi are prolific producers of natural products, many of which have biological activities that aid the fungi in colonizing and killing plants^{1,2,18}. Therefore, fungal natural products represent a promising source of potential leads for herbicides. The abundance of sequenced fungal genomes enables genome mining of new

natural products with novel biological activities^{4,6}. Although no natural-product inhibitors of DHAD are known to date, we reason that a fungal natural product with this property might exist, given the indispensable role of BCAA biosynthesis in plants¹⁰.

To identify natural-product biosynthetic gene clusters that may encode a DHAD inhibitor, we hypothesized that such a cluster must contain an additional copy of DHAD that is insensitive to the inhibitor, thereby providing the required self-resistance for the producing organism to survive. Genes encoding a self-resistance enzyme are frequently found in microbial natural-product gene clusters, as highlighted by the presence of an insensitive copy of HMG-CoA reductase (HMGR) and inosine monophosphate dehydrogenase (IMPDH) in the gene clusters of lovastatin (that targets HMGR) and mycophenolic acid (that targets IMPDH), respectively^{19,20} (Extended Data Fig. 1c). This phenomenon has been used to predict molecular targets of natural products, as well as to identify gene clusters of natural products of known activities^{5,7,9}.

To identify possible self-resistance enzymes, we scanned sequenced fungal genomes to search for co-localization of genes encoding DHAD with core biosynthetic enzymes, such as terpene cyclases and polyketide synthases among others^{21,22}. We identified a well-conserved set of four genes across multiple fungal genomes (Fig. 1b), including the common soil fungus Aspergillus terreus that is best known for producing lovastatin. The conserved gene clusters include genes that encode a sesquiterpene cyclase homologue (astA), two cytochrome P450 genes (astB and astC) and a homologue of DHAD (astD). Genes outside of this cluster are not conserved across the identified genomes and are hence unlikely to be involved in the biosynthesis of natural products. AstD is the second copy of DHAD encoded in the genome, and is approximately 70% similar to the housekeeping copy that is well-conserved across fungi (Supplementary Fig. 2). Therefore, AstD is potentially a self-resistance enzyme that confers resistance to the encoded natural product. As with a majority of biosynthetic gene clusters in sequenced fungal genomes, the ast cluster has not been associated with the production of a known natural product⁴.

To identify the natural product encoded by the *ast* cluster, we heterologously expressed *astA*, *astB* and *astC* genes in the host *Saccharomyces cerevisiae* RC01²³. New compounds that emerged were purified and their structures were elucidated using NMR spectroscopy (Supplementary Fig. 3, Supplementary Table 5). RC01 cells expressing only *astA* produced a new sesquiterpene (1), which was confirmed to be (—)-daucane (Supplementary Fig. 4). RC01 cells expressing both *astA* and *astB* led to the biosynthesis of a new product that was determined structurally to be the α -epoxy carboxylate (2) (Fig. 1c). When *astA*, *astB* and *astC* were expressed together, a new compound (3) became the dominant product (approximately 20 mg l⁻¹). Full structural determination revealed 3 to be the tricyclic aspterric acid, which is a previously isolated compound²⁴ (Fig. 1c). The biosynthetic pathway for aspterric

¹Department of Chemical and Biomolecular Engineering, University of California Los Angeles, Los Angeles, CA, USA. ²Department of Molecular, Cell, and Developmental Biology and Howard Hughes Medical Institute, University of California Los Angeles, CA, USA. ³State Key Laboratory of Bio-organic and Natural Products Chemistry, Center for Excellence in Molecular Synthesis, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, China. ⁴Department of Chemistry, Shanghai Normal University, Shanghai, China. ⁵Laboratory of Physical Chemistry of Polymers and Membranes, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. ⁶State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Physiology and Biophysics, School of Life Sciences, Fudan University, Shanghai, China. ⁷Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA, USA. ⁸These authors contributed equally: Yan Yan, Qikun Liu, Xin Zang. *e-mail: jiahai@mail.sioc.ac.cn; jacobsen@ucla.edu; yitang@ucla.edu;

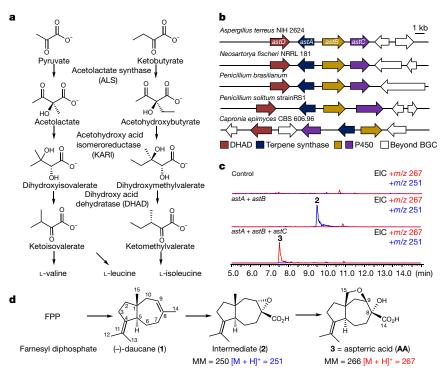


Fig. 1 | Genome mining of a DHAD inhibitor and biosynthesis of aspterric acid. a, Valine, leucine and isoleucine are produced by two parallel pathways using three enzymatic steps: ALS, KARI and DHAD. b, A 17-kb biosynthetic gene cluster (BGC) from *A. terreus* containing four open reading frames (ORFs), which are also conserved among several fungal species. *astA* has sequence homology to sesquiterpene cyclase; *astB* and *astC* are predicted to be P450 monooxygenases; *astD* is predicted to encode a DHAD and is proposed to confer self-resistance in the presence of the natural product produced in the cluster. c, High-performance liquid chromatography–mass spectrometry (HPLC–MS) traces of metabolites

acid is therefore concise: after cyclization of farnesyl diphosphate by AstA to create the carbon skeleton in 1, AstB catalyses oxidation of 1 to yield the epoxide 2. Further oxidation by AstC at carbon 15 yields an alcohol, which can undergo intramolecular epoxide opening to create aspterric acid (Fig. 1d).

Upon its initial discovery, aspterric acid was shown to have inhibitory activity towards Arabidopsis thaliana, however, the mode of action was not known²⁵. Our resistance-gene-directed approach led to rediscovery of this compound with DHAD as a potential target. We first confirmed that aspterric acid is able to potently inhibit A. thaliana growth in an agar-based assay (Fig. 2a, Supplementary Fig. 5). Aspterric acid was also an effective inhibitor of root development and plant growth when applied to a representative monocot (Zea mays) and dicot (Solanum lycopersicum) (Fig. 2b). To test whether aspterric acid indeed targets DHAD, we expressed and purified housekeeping DHAD from both A. terreus (XP_001208445.1, AteDHAD) and A. thaliana (AT3G23940, AthDHAD), as well as the putative self-resistance enzyme AstD (Supplementary Fig. 6). Both housekeeping DHAD enzymes converted dihydroxyisovalerate to ketoisovalerate (AthDHAD: $k_{cat} = 1.2 \text{ s}^{-1}$, $K_{\rm M} = 5.7$ mM) as expected. The enzyme activities, however, were inhibited in the presence of aspterric acid (Extended Data Fig. 2). The half-maximal inhibitory concentration (IC₅₀) values of aspterric acid towards AteDHAD and AthDHAD were 0.31 μM and 0.50 μM, respectively, at an enzyme concentration of 0.50 μM (Extended Data Fig. 3). Aspterric acid was further determined to be a competitive inhibitor of AthDHAD with a $K_i = 0.30 \,\mu\text{M}$ (Extended Data Fig. 3). Aspterric acid displayed no significant cytotoxicity towards human cell lines up to $500\,\mu\text{M}$ concentration, consistent with the lack of DHAD in mammalian cells (Supplementary Fig. 7).

AstD catalyses the identical β -dehydration reaction as DHAD, albeit with a significantly slower turnover rate ($k_{cat} = 0.03 \text{ s}^{-1}$, $K_{M} = 5.4 \text{ mM}$).

produced from *S. cerevisiae* RC01 expressing the different *ast* genes under the P_{ADH2} promoter control. Control, *S. cerevisiae* without expression plasmids; *S. cerevisiae* transformed with plasmids expressing *astA* and *astB*, which produces 2; and *S. cerevisiae* transformed with plasmids expressing *astA*–*C*, which produces aspterric acid at a titre of 20 mg l⁻¹. The experiments were repeated independently with similar results three times. EIC, extracted ion chromatogram. **d**, Proposed biosynthetic pathway of aspterric acid. AstA cyclizes farnesyl diphosphate (FPP) into 1 and the P450 enzymes AstB and AstC then sequentially transform 1 into 2 and aspterric acid (3, AA), respectively. MM, molecular mass.

However, the enzyme was not inhibited by aspterric acid, even at the solubility limit of 8 mM (Extended Data Fig. 3). To determine if AstD can confer resistance to aspterric acid-sensitive strains, we developed a yeast-based assay. The genome copy of DHAD encoded by *ILV3* was first deleted from the S. cerevisiae strain DHY $\Delta URA3$, which resulted in an auxotroph that requires exogenous addition of Ile, Leu and Val to grow. We introduced either the gene encoding AteDHAD or astD episomally into the ILV3 knockout strain and found that both genes enabled the cells to grow in the absence of the three BCAAs (Extended Data Fig. 4). However, yeast cells expressing AteDHAD were approximately 100 times more sensitive to aspterric acid (IC₅₀ of 2 μM) compared to yeast expressing AstD (IC $_{50}$ of 200 $\mu M)$ (Fig. 2c). Collectively, the biochemical and genetic assays validated that aspterric acid is, to our knowledge, the first natural-product inhibitor of fungal and plant DHAD, and that AstD serves as the self-resistance enzyme in the ast biosynthetic gene cluster.

The (R)- α -hydroxyacid and (R)-configured β -ether oxygen moieties in aspterric acid mimic the (2R,3R)-dihydroxy groups present in natural substrates such as dihydroxyisovalerate. The β -ether oxygen in aspterric acid is in a position to coordinate to the 2Fe–2S cluster that is a required cofactor in both fungal and plant DHAD^{16,17}. To understand the potential mechanism of action of aspterric acid, we determined the crystal structure (2.11~Å) of AthDHAD in complex with the 2Fe–2S cluster (holo-AthDHAD) (Fig. 2d, Extended Data Fig. 5, Extended Data Table 1). We identified a binding chamber at the homodimer interface, similar to that found in the holo bacterial L-arabinonate dehydratase²⁶ (Fig. 2d). The interior of the chamber is positively charged (2Fe–2S and Mg^{2+}) whereas the entrance is lined with hydrophobic residues. The modelled binding mode of α , β -dihydroxyisovalerate and aspterric acid predicted by computational docking are shown in Fig. 2e. The pocket is sufficiently spacious to accommodate the bulkier aspterric

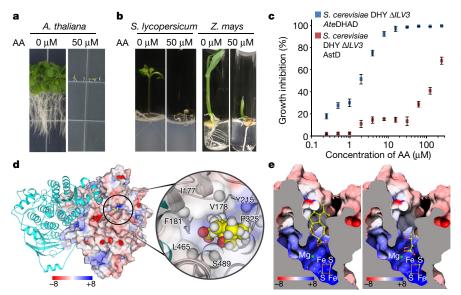


Fig. 2 | Aspterric acid is a plant growth inhibitor. a, Two-week-old A. thaliana growing on Murashige and Skoog basal medium containing no aspterric acid (left) or 50 μ M aspterric acid (right). The picture shown is representative of three replicates. b, Same as in a, except for two-week-old dicotyledon S. lycopersicum and monocotyledon Z. mays. The picture shown is representative of two replicates. c, Verification of the self-resistance function of AstD. Growth-inhibition curve of aspterric acid on S. cerevisiae $\Delta ILV3$ strains expressing fungal housekeeping AteDHAD (blue) or AstD (red) in isoleucine, leucine and valine (ILV) dropout medium. Data are mean \pm s.d. from three biologically independent experiments. d, Crystal structure of dimeric holo-AthDHAD containing the cofactor 2Fe–2S cluster and a Mg²+ ion with the docked aspterric acid in the

active site. One of the *Ath*DHAD monomers is show in cyan, whereas the other one is shown in electrostatic surface representation. The docked aspterric acid is shown inset as a spaced-filled model. The hydrophobic portions of aspterric acid are surrounded by several hydrophobic residues (white spheres) from both monomers. **e**, Cross-section electrostatic map of modelled holo-*Ath*DHAD in the binding site. Red surface map, the normalized negatively charged regions; blue surface map, the normalized positively charged regions; white surface map, the hydrophobic regions. The docked aspterric acid in the active site of *Ath*DHAD is shown on the left, and the docked native substrate dihydroxyisovalerate is shown on the right. The docking studies suggest the hydrophobic entrance to the reaction chamber preferentially binds the bulkier, tricyclic aspterric acid.

acid, and provide stronger hydrophobic interactions than the native substrate with a 5.3 ± 0.3 kcal mol $^{-1}$ gain in binding energy (Fig. 2e). On the basis of the holo-AthDHAD structure, we constructed a homology model of AstD to determine the potential mechanism of resistance (Extended Data Figs. 5, 6). Comparison of AthDHAD and the modelled AstD structures shows that although most of the residues in the catalytic chamber are conserved, the hydrophobic region at the entrance to the reactive chamber in AstD is more constricted as a result of two amino acid substitutions (V496L and I177L). Narrowing of the entrance could therefore sterically exclude the bulkier aspterric acid from binding in the active site, whereas the smaller, natural substrates are still able to enter the chamber.

To explore the potential of aspterric acid as an herbicide, we performed spray treatment of A. thaliana with aspterric acid. We added aspterric acid into a commercial glufosinate formulation known as Finale at a final concentration of 250 μ M^{27,28}. We then sprayed aspterric acid solution onto glufosinate-resistant A. thaliana. Finale alone had no observable inhibitory effects on plant growth, but adding aspterric acid severely inhibited plant growth (Extended Data Fig. 7). In addition, A. thaliana plants treated with aspterric acid before flowering failed to form normal pollen, which has also been observed previously²⁵. We found that the pistil of treated plants could still be successfully pollinated using healthy pollen from the untreated A. thaliana, indicating that aspterric acid preferentially affects pollen but not egg formation (Extended Data Figs. 8, 9). This effect was also observed with a lower concentration of aspterric acid (100 μ M). Thus, in addition to its herbicidal properties, aspterric acid could potentially be used as a chemical hybridization agent for hybrid seed production²⁹.

We next investigated whether plants expressing *astD* are resistant to aspterric acid. This was motivated by the successful combination of glyphosate and genetically modified crops that are selectively resistant to glyphosate (Roundup Ready)³⁰. The *A. terreus astD* gene was codon optimized and the N terminus was fused to a chloroplast localization signal derived from *Ath*DHAD. Wild-type or *astD* transgene-expressing

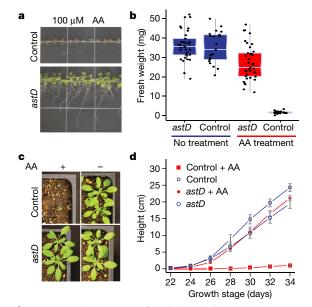


Fig. 3 | Aspterric acid-resistance of *Arabidopsis* plants expressing *astD* transgenes. a, Phenotype of ten-day-old *A. thaliana* with (lower) and without (upper) the *astD* transgene growing on medium containing $100 \, \mu M$ aspterric acid. Control plants were transformed with a vector that carries the glufosinate ammonium selection marker but no *astD* transgene. The picture shown is representative of three replicates. **b**, Fresh weight of three-week-old *Arabidopsis* seedlings growing on medium with (red box) and without (blue box) $100 \, \mu M$ aspterric acid. Box plots show the median and whiskers extend to the first and third quartiles, with the individual data points from 21 biologically independent experiments overlaid. **c**, Glufosinate-resistant *Arabidopsis* with (lower) and without (upper) *astD* transgene growing in soil were sprayed with glufosinate ammonium with (left) and without (right) $250 \, \mu M$ aspterric acid. **d**, Quantification of the height of *Arabidopsis* treated as in **c**. Data are mean \pm s.d. from 12 biologically independent experiments.

A. thaliana was then grown on medium that contained 100 μ M aspterric acid. In the presence of aspterric acid, the growth of wild-type plants was strongly inhibited, and arrested at the cotyledon stage (Fig. 3a). By contrast, the growth of astD transgenic plants was relatively unaffected by aspterric acid, as indicated by the normally expanded rosette leaves, elongated roots and whole-plant fresh weight (Fig. 3a, b). The expression of AstD was verified by western blot (Supplementary Fig. 8). A spray assay was also performed using T2 astD transgenic A. thaliana plants, which showed no observable growth defects under such treatment (Fig. 3c). By contrast, the control plants carrying the empty vector showed a strong growth inhibitory phenotype when treated with aspterric acid (Fig. 3c). Quantitative measurements of plant height showed that AstD effectively confers aspterric acid resistance to A. thaliana (Fig. 3d).

In summary, resistance-gene-directed discovery of natural products in the fungus *A. terreus* led to the discovery of a natural herbicide aspterric acid and the determination of its mode of action. In addition, introducing *astD* as a transgene or editing the sequence of the plant DHAD endogenous gene could be used to create aspterric acid-resistant crops. We suggest that aspterric acid is a promising lead for development as a broad spectrum commercial herbicide.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0319-4

Received: 21 November 2017; Accepted: 22 May 2018; Published online 11 July 2018.

- Dayan, F. E. & Duke, S. O. Natural compounds as next-generation herbicides. Plant Physiol. 166, 1090–1105 (2014).
- Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. J. Nat. Prod. 79, 629–661 (2016).
- Clardy, J. & Walsh, C. Lessons from natural molecules. Nature 432, 829–837 (2004).
- Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes—a review. Nat. Prod. Rep. 33, 988–1005 (2016).
- Clevenger, K. D. et al. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat. Chem. Biol. 13, 895–901 (2017).
- Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* 13, 509–523 (2015)
- Tang, X. et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem. Biol. 10, 2841–2849 (2015).
- Alanjary, M. et al. The antibiotic resistant target seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. Nucleic Acids Res. 45, W42–W48 (2017).
- Yeh, H.-H. et al. Resistance gene-guided genome mining: serial promoter exchanges in Aspergillus nidulans reveal the biosynthetic pathway for fellutamide B, a proteasome inhibitor. ACS Chem. Biol. 11, 2275–2284 (2016).
- Amorim Franco, T. M. & Blanchard, J. S. Bacterial branched-chain amino acid biosynthesis: structures, mechanisms, and drugability. *Biochemistry* 56, 5849–5865 (2017).
- Heap, I. Global perspective of herbicide-resistant weeds. Pest Manag. Sci. 70, 1306–1315 (2014).
- Soltani, N. et al. Potential corn yield losses from weeds in North America. Weed Technol. 30, 979–984 (2016).
- Swanton, C. J., Harker, K. N. & Anderson, R. L. Crop losses due to weeds in Canada. Weed Technol. 7, 537–542 (1993).
- Gianessi, L. P. The increasing importance of herbicides in worldwide crop production. Pest Manag. Sci. 69, 1099–1105 (2013).
- Flint, D. H. & Emptage, M. H. Dihydroxy acid dehydratase from spinach contains a [2Fe–2S] cluster. J. Biol. Chem. 263, 3558–3564 (1988).

- Flint, D. H., Emptage, M. H., Finnegan, M. G., Fu, W. & Johnson, M. K. The role and properties of the iron–sulfur cluster in *Escherichia coli* dihydroxy-acid dehydratase. *J. Biol. Chem.* 268, 14732–14742 (1993).
- Flint, D. H. & Nudelman, A. Studies on the active site of dihydroxy-acid dehydratase. *Bioorg. Chem.* 21, 367–385 (1993).
- Pusztahelyi, T., Holb, I. J. & Pócsi, I. Secondary metabolites in fungus–plant interactions. Front. Plant Sci. 6, 573 (2015).
- Kennedy, J. et al. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. Science 284, 1368–1372 (1999).
- Regueira, T. B. et al. Molecular basis for mycophenolic acid biosynthesis in Penicillium brevicompactum. Appl. Environ. Microbiol. 77, 3035–3043 (2011).
- Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* 106, 3468–3496 (2006).
- Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. 106, 3412–3442 (2006).
- Tang, M.-C. et al. Discovery of unclustered fungal indole diterpene biosynthetic pathways through combinatorial pathway reassembly in engineered yeast. J. Am. Chem. Soc. 137, 13724–13727 (2015).
- Yoshisuke Tsuda et al. Aspterric acid, a new sesquiterpenoid of the carotane group, a metabolite from Aspergillus terreus IFO-6123. X-ray crystal and molecular structure of its p-bromobenzoate. J. Chem. Soc. Chem. Commun. 0, 160–161 (1978).
- Shimada, A. et al. Aspterric acid and 6-hydroxymellein, inhibitors of pollen development in Arabidopsis thaliana, produced by Aspergillus terreus. Z. Naturforsch. C 57, 459–464 (2002).
- Rahman, M. M. et al. The crystal structure of a bacterial l-arabinonate dehydratase contains a [2Fe–2S] cluster. ACS Chem. Biol. 12, 1919–1927 (2017).
- Kirkwood, R. C. Use and mode of action of adjuvants for herbicides: a review of some current work. Pestic. Sci. 38, 93–102 (1993).
- Hoerlein, G. Glufosinate (phosphinothricin), a natural amino acid with unexpected herbicidal properties. Rev. Environ. Contam. Toxicol. 138, 73–145 (1994).
- McRae, D. H. in *Plant Breeding Reviews* (ed. Janick, J.) 169–191 (John Wiley & Sons, New Jersey, 1985).
- Benbrook, C. M. Trends in glyphosate herbicide use in the United States and globally. *Environ. Sci. Eur.* 28, 3–19 (2016).

Acknowledgements This work was supported by the NIH (1DP1GM106413 and 1R35GM118056) to Y.T. and CAS (XDB20000000) to J.Z. S.E.J. is an Investigator of the Howard Hughes Medical Institute. Q.L. is supported by the NIH (F32) Postdoctoral Fellowship. We thank Stanford Genome Technology Center for the S. cerevisiae DHY $\Delta URA3$ strain. The diffraction data of holo-AthDHAD was collected at beamline BL19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). The molecular modelling was performed at the Interdisciplinary Centre for Mathematical and Computational Modeling in Warsaw (GB70-3 & GB71-3). We thank W. Huang, L. Wu and R. Cheng for technical help with protein purification and crystallization.

Reviewer information *Nature* thanks F. Dayan, E. Sattely and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.Y., Q.L., J.Z., S.E.J. and Y.T. developed the hypothesis and designed the study. Y.Y. performed all in vivo and in vitro experiments. Q.L., Y.Y. and C.N. performed plant experiments. X.Z. and J.Z. performed crystallography experiments. J.G. determined the protein structure. S.Y. performed computational experiments. Y.Y. and U.B.-E. performed yeast experiments. Y.Y., Q.L., J.Z., S.E.J. and Y.T. prepared the manuscript.

Competing interests A PCT patent application related to this manuscript has been filed by UCLA.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0319-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0319-4.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.Z., S.E.J. and Y.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Materials. Biological reagents, chemicals, media and enzymes were purchased from standard commercial sources unless stated. The plant, fungal, yeast and bacterial strains, plasmids and primers used in this study are summarized in Supplementary Tables 3, 4 and 5. DNA and RNA manipulations were carried out using Zymo ZR Fungal/Bacterial DNA Microprep kit and Invitrogen Ribopure kit respectively. DNA sequencing was performed at Laragen. The primers and codon-optimized gblocks were synthesized by IDT.

Expression of ast genes in Aspergillus nidulans for cDNA isolation. Plasmids pYTU, pYTP, pYTR were digested with PacI and SwaI and used as vectors to insert genes³¹. A gpda promoter was generated by PCR amplification using primers Gpda-pYTU-F and Gpda-R with pYTR serving as template. Genes to be expressed were amplified using PCR with the genomic DNA of A. terreus NIH2624 as a template. A 4.5-kb fragment obtained using primers AstD-pYTU-recomb-F and AstA-pYTU-recomb-R was cloned into pYTU together with a gpda promoter by yeast homologous recombination to obtain pAstD+AstA-pYTU. Yeast transformation was performed using Frozen-EZ Yeast Transformation II Kit (Zymo Research). A 2.4-kb fragment obtained using primers AstB-pYTR-recomb-F and AstB-pYTR was cloned into pYTR by yeast homologous recombination to obtain pAstB-pYTR. Similarly, a 2.3-kb fragment obtained using primers AstC-pYTP-recomb-F and AstC-pYTP-recomb-R was cloned into pYTP by yeast homologous recombination to obtain pAstC-pYTP.

All three plasmids (pAstD+AstA-pYTU, pAstB-pYTR and pAstC-pYTP) were transformed into *A. nidulans* according to standard protocols to result in the *A. nidulans* strain TY01³¹. TY01 was cultured in liquid CD-ST medium (20 g l⁻¹ starch, 20 g l⁻¹ peptone, 50 ml l⁻¹ nitrate salts and 1 ml l⁻¹ trace elements) at 28 °C for 3 days. Total RNA of TY01 was extracted with the Invitrogen Ribopure kit, and total cDNA of TY01 was obtained using the SuperScript III reverse transcriptase kit (Thermo Fisher Scientific). The cDNA fragment of *astA* was PCR amplified using primers AstA-xw55-recomb-F and AstA-xw55-recomb-R. The cDNA fragment of *astB* was PCR amplified using primers AstB-xw06-recomb-F and AstB-xw06-recomb-R. The cDNA fragment of *astC* was PCR amplified using primers AstC-xw02-recomb-F. and AstC-xw02-recomb-F. The cDNA fragment of *astD* was PCR amplified using primers AstD-pXP318-F. All the introns were confirmed to be correctly removed by sequencing.

Construction of S. cerevisiae strains. Plasmid pXW55 (*URA3* marker) digested with *NdeI* and *PmeI* was used to introduce the *astA* gene²³ into *S. cerevisiae* RC01. A 1.3-kb fragment containing *astA* obtained from PCR using primers AstA-xw55-recomb-F and AstA-xw55-recomb-R was cloned into pXW55 using yeast homologous recombination to produce pAstA-xw55. The plasmid pAstA-xw55 was then transformed into *S. cerevisiae* RC01 to generate strain TY02²³.

Plasmid pXW06 (*TRP1* marker) digested with *NdeI* and *PmeI* was used to introduce the *astB* gene²³ *S. cerevisiae* RC01. A 1.6-kb fragment containing *astB* obtained from PCR using primers AstB-xw06-recomb-F and AstB-xw06-recomb-R were cloned into pXW06 using yeast homologous recombination to produce pAstB-xw06. The plasmid pAstB-xw06 was then transformed into TY02 to generate strain TY03.

Plasmid pXW06 (*LEU2* marker) digested with *NdeI* and *PmeI* was used to introduce the *astC* gene²³ *S. cerevisiae* RC01. A 1.6-kb fragment containing *astC* obtained from PCR using primers AstC-xw02-recomb-F and AstC-xw02-recomb-R were cloned into pXW02 using yeast homologous recombination to produce pAstC-xw02. The plasmid pAstC-xw02 was then transformed into TY03 to generate strain TY04.

The URA3 gene was inserted into the ILV3 locus of S. cerevisiae DHY $\Delta URA3$ strain to generate UB01. A 879-bp homologous-recombination donor fragment with 35–40 bp homologous regions flanking the ILV3 ORF was amplified using primers ILV3p-URA3-F and ILV3t-URA3-R using yeast gDNA as a template. The PCR product was gel purified and transformed into S. cerevisiae DHY $\Delta URA3$, and selected on uracil dropout medium to give UB01. The resulting strain was subjected to verification using colony PCR with primers ILV3KO-ck-F and ILV3KO-ck-R and the amplified fragment was confirmed with sequencing.

The URA3 gene inserted into the ILV3 locus of S. cerevisiae DHY $\Delta URA3$ was deleted from UB01 using homologous recombination to generate UB02. A 150-bp homologous-recombination donor fragment with 75-bp homologous regions flanking the ILV3 ORF was amplified using primers ILV3KO-F and ILV3KO-R, gel purified, transformed into UB01, and counter-selected on 5-fluoroorotic acid (5-FoA)-containing medium to give UB02. The resulting strain was subjected to verification using colony PCR with primers ILV3KO-ck-F and ILV3KO-ck-R and the amplified fragment was confirmed with sequencing.

The empty plasmid pXP318 (URA3 marker) was transformed into UB02 to generate TY05³².

Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce the gene encoding *Ate*DHAD³² into the plasmid UB02. The cDNA of *A. terreus* NIH 2624 served as the template for PCR amplification. A 1.7-kb fragment obtained using primers *Ate*DHAD-pXP318-F and *Ate*DHAD-pXP318-R were cloned into pXP318 using yeast homologous recombination to produce *Ate*DHAD-pXP318. Then, *Ate*DHAD-pXP318 was transformed into UB02 to generate TY06. *Ate*DHAD was driven by a constitutive *TEF1* promoter.

Plasmid pXP318 digested with *Spe*I and *Xho*I was used as vector to introduce the *astD* gene³² into the plasmid UB02. The cDNA isolated from TY01 served as the template for PCR amplification. A 1.8-kb fragment obtained using primers AstD-pXP318-F and AstD-pXP318-R were cloned into pXP318 using yeast homologous recombination to make AstD-pXP318. A Flag tag was also added to the N-terminal of AstD. AstD-pXP318 was then transformed into UB02 to generate TY07. AstD was driven by the constitutive *TEF1* promoter.

Fermentation and compound analyses and isolation. A seed culture of *S. cerevisiae* strain was grown in 40 ml of synthetic dropout medium for 2 days at $28\,^{\circ}\text{C}$, 250 r.p.m. Fermentation of the yeast was carried out using YPD (yeast extract $10\,\text{g l}^{-1}$, peptone $20\,\text{g l}^{-1}$) supplemented with 2% dextrose for 3 days at $28\,^{\circ}\text{C}$, $250\,\text{r.p.m.}$

HPLC–MS analyses were performed using a Shimadzu 2020 EVLC-MS (Phenomenex Luna, 5μ , 2.0×100 mm, C-18 column) with positive and negative mode electrospray ionization. The elution method was a linear gradient of 5–95% (v/v) acetonitrile/water over 15 min, and then 95% (v/v) acetonitrile/water for 3 min with a flow rate of 0.3 ml min $^{-1}$. The HPLC buffers were supplemented with 0.05% formic acid (v/v). HPLC purifications were performed using a Shimadzu Prominence HPLC (Phenomenex Kinetex, 5μ , 10.0×250 mm, C-18 column). The elution method was a linear gradient of 65–100% (v/v) acetonitrile/water in 25 min, with a flow rate of 2.5 ml min $^{-1}$. Gas chromatography—mass spectrometry (GC–MS) analyses were performed using Agilent Technologies GC–MS 6890/5973 equipped with a DB-FFAP column. An inlet temperature of 240 °C and constant pressure of 4.2 psi were used. The oven temperature was initially set at 60 °C, then ramped up at 10 °C min $^{-1}$ for 20 min and finally held at 240 °C for 5 min.

To isolate compound 1, the fermentation broth of TY02 was centrifuged (5,180g, 10 min), and the cell pellet was collected and soaked in acetone. The organic phase was dried over sodium sulfate, concentrated to oil form and subjected to silica column purification with hexane. To isolate compound 2, the fermentation broth of TY03 was centrifuged (5,180g, 10 min), and the supernatant was extracted three times with ethyl acetate. The organic phase was dried over sodium sulfate, concentrated to oil form, and then and subjected to HPLC purification. To isolate aspterric acid, the fermentation broth of TY04 was centrifuged (5,180g, 10 min), and supernatant was extracted three times with ethyl acetate. The organic phase was dried over sodium sulfate, concentrated to oil form, and subjected to HPLC purification.

Structure determination of compounds. Compound 1, a colourless oil that readily dissolved in hexane and chloroform, had a molecular formula $C_{15}H_{24}$, as deduced from electron ionization—mass spectrometry (EI–MS) [M]+ m/z 204, and showed $[\alpha]_D^{22} = -30^\circ$ (n-hexane; c = 0.1). GC–MS 70 eV, m/z (relative intensity): 204 [M]+ (42), 189 (5), 161 (35), 136 (100), 133 (10), 121 (70), 119 (25), 107 (20), 105 (27), 93 (21), 91 (26), 79 (13), 77 (15), 69 (20), 55 (12), 43 (12), 41 (13), 38 (21); 1 H NMR (500 MHz, CDCl₃): δ (p.p.m.) 5.37 (1H, m), 2.20–2.10 (5H, m), 2.10–2.00 (2H, m), 1.95 (1H, d, 15.3), 1.75 (3H, s), 1.71 (3H, q, 1.7), 1.61 (3H, br s), 1.44 (1H, dd, 11.4, 7.2), 1.36 (1H, m), 1.31 (1H, dd, 11.3, 2.6), 0.73 (3H, s); 13 C NMR (125 MHz, CDCl₃): δ 138.4, 138.3, 122.4, 122.2, 57.4, 42.6, 41.4, 40.3, 34.5, 29.6, 27.3, 25.0, 23.3, 20.6, 19.2. Both of the NMR and mass spectrometry spectra are identical to a known compound (+)-daucane, however, the optical rotation is opposite which led to the assignment of 1 to be (–)-daucane³³.

Compound **2**, a colourless oil that readily dissolved in ethyl acetate and chloroform, had a molecular formula $C_{15}H_{22}O_3$, as deduced from liquid chromatographymass spectrometry (LC–MS) [M+H]⁺ m/z 251, [M – H]⁻ m/z 249. ¹H NMR (500 MHz, CDCl₃): δ 8.09 (1H, brs), 3.25 (1H, t, 7.4), 2.71 (1H, dd, 14.6, 6.5), 2.48 (1H, dd, 14.8, 6.3), 2.36 (1H, dd, 14.0, 6.6), 2.26 (1H, m), 2.15 (1H, dd, 16.3, 8.9), 2.08 (1H, d, 12.0), 1.84 (1H, q, 13.1), 1.73 (3H, d, 2.3), 1.59 (3H, d, 2.2), 1.48–1.35 (3H, m), 1.31 (1H, td, 11.5, 9.0), 0.86 (3H, s). ¹³C NMR (125 MHz, CDCl₃): δ 176.0, 135.8, 123.2, 60.1, 59.8, 59.4, 44.1, 40.5, 38.8, 30.6, 29.3, 24.9, 23.8, 20.6, 17.8.

Compound **3**, a colourless oil that readily dissolved in acetone and chloroform, had a molecular formula $C_{15}H_{22}O_4$, as deduced from LC–MS $[M+H]^+$ m/z 267, $[M-H]^-$ m/z 265. 1H NMR (500 MHz, CDCl₃): δ 4.29 (1H, d, 8.5), 3.92 (1H, d, 8.3), 3.48 (1H, d, 8.3), 2.42 (1H, dd, 14.9, 7.3), 2.37–2.28 (2H, m), 2.25 (1H, dd, 13.0, 4.4), 2.20–2.17 (1H, m), 2.12 (1H, d, 13.4), 2.01 (1H, m), 1.80–1.65 (2H, m), 1.71 (3H, s), 1.64–1.54 (1H,m), 1.60 (3H, s), 1.50 (1H, m); 13 C NMR (125 MHz, CDCl₃): δ 178.2, 134.5, 125.2, 82.9, 76.3, 75.6, 55.4, 53.0, 36.6, 36.2, 33.8, 32.2, 23.6, 23.4, 20.9. Compound **3** is identical to aspterric acid as reported 24,25 .

Protein expression, purification and biochemical assay. To express and purify *Ath*DHAD, primers *Ath*DHAD-pET-F and *Ath*DHAD-pET-R were used to amplify a 1.7-kb DNA fragment containing *Ath*DHAD (AT3G23940). The PCR

product was cloned into pET28a using NheI and NotI restriction sites. The resulting plasmid AthDHAD-pET was transformed into E.coli BL21 (DE3) to give TY08. To express and purify AteDHAD (XP_001208445.1), primers AteDHAD-pET-F and AteDHAD-pET-R were used to amplify a 1.6-kb DNA fragment containing AteDHAD. The PCR product was cloned into pET28a using NdeI and NotI restriction sites. The resulting plasmid AteDHAD-pET was transformed into E. coli BL21 (DE3) to obtain TY09. To express and purify AstD (XP_001213593.1), primers AstD-pET-F and AstD-pET-R were used to amplify a 1.6-kb DNA fragment containing astD. The PCR product was cloned into pET28a using NdeI and NotI restriction sites. The resulting plasmid AstD-pET was transformed into E. coli BL21 (DE3) to obtain TY10. All DHADs with a fused 6 × His-tag with a molecular mass of ~62 kDa were expressed at 16 °C 220 rpm for 20 h after 100 μM isopropyl β -D-1-thiogalactopyranoside IPTG induction (IPTG was added when $\mathrm{OD}_{600\,\mathrm{nm}} = 0.8$). Cells from a 1-l culture were then collected by centrifugation at 5,180g at 4°C. Cell pellet was resuspended in 15 ml Buffer A10 (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 8% glycerol, 10 mM imidazole). The cells were lysed by sonication, and the insoluble material was sedimented by centrifugation at 35,267g at 4°C. The protein supernatant was then incubated with 3 ml Ni-NTA for 4 h with slow, constant rotation at 4°C. Subsequently the Ni-NTA resin was washed with ten column volumes of Buffer A50 (Buffer A with 50 mM imidazole). For elution of the target protein, the Ni-NTA resin was incubated for 10 min with 6 ml Buffer A300 (Buffer A with 300 mM imidazole). The supernatant from the elution step was then analysed by SDS-PAGE together with the supernatants from the other purification steps. The elution fraction containing the recombinant protein was buffer exchanged into storage buffer (50 mM Tris-HCl pH 7.2, 50 mM NaCl, 10 mM MgCl₂, 10% glycerol, 5 mM DTT, 5 mM GSH).

In vitro activity assays were carried out in 50 μl reaction mixture containing storage buffer, 10 mM (\pm)-sodium α,β -dihydroxyisovalerate hydrate (4) and 0.5 μM of purified DHAD enzyme. The reaction was initiated by adding the enzyme. After 0.5 h incubation at 30 °C, the reactions were stopped by adding an equal volume of ethanol. Approximately 0.1 volumes of 100 mM phenylhydrazine (PHH) was added to derivatize the product 3-methyl-2-oxo-butanoic acid (5) into 6 at room temperature for 30 min. 20 μl of the reaction mixture was used for the LC–MS analysis. The area of the HPLC peak with UV absorption at 350 nm were used to quantify the amount of 6. (Extended Data Fig. 2).

The inhibition percentage of aspterric acid on DHADs was determined using in vitro biochemical assays and calculated with following equation:

$$inhibition \ percentage = 1 - \frac{initial \ reaction \ rate \ with \ aspterric \ acid}{initial \ reaction \ rate \ without \ aspterric \ acid}$$

Growth inhibition assay of S. cerevisiae on plates or in the tubes. S. cerevisiae was grown in isoleucine, leucine and valine (ILV) dropout medium (20 g l⁻¹ glucose, $0.67~{\rm g}~{\rm l}^{-1}$ Difco Yeast Nitrogen Base without amino acids, $18~{\rm mg}~{\rm l}^{-1}$ adenine, arginine 76 mg l^{-1} , asparagine 76 mg l^{-1} , aspartic acid 76 mg l^{-1} , glutamic acid 76 mg l^{-1} , histidine 76 mg l^{-1} , lysine 76 mg l^{-1} , methionine 76 mg l^{-1} , phenylalanine 76 mg l^{-1} , serine 76 mg l^{-1} , threonine 76 mg l^{-1} , tryptophan 76 mg l^{-1} , tyrosine 76 mg l^{-1}) to test growth inhibition of aspterric acid on *S. cerevisiae*, cells were incubated at 28 $^{\circ}\text{C}$ until OD $_{600\,\mathrm{nm}}$ of the control strain without aspterric acid treatment reached about 0.8. The ratio of yeast $\mathrm{OD}_{600\,\mathrm{nm}}$ in medium with aspterric acid treatment to yeast OD₆₀₀ in medium without aspterric acid was calculated as the percentage of growth inhibition. The inhibition curve was plotted as the percentage of inhibition versus aspterric acid concentrations. To further prove aspterric acid affects BCAA biosynthesis, isoleucine, leucine and valine were also complemented to the medium with or without treatment with aspterric acid. The growth curves of TY05, TY06 and TY07 were also plotted in Extended Data Fig. 4. The $\mathrm{OD}_{600\,\mathrm{nm}}$ was recorded for every 20 min over a total of 50 h. The growth inhibition percentage of aspterric acid on S. cerevisiae strain is calculated by dividing the cell density $(OD_{600 \text{ nm}})$ of the aspterric acid-treated strain to the corresponding untreated strains when $OD_{600\,nm}$ reaches approximately 0.8 using the following equation:

growth inhibition percentage =
$$1 - \frac{\mathrm{OD}_{600\;\mathrm{nm}}}{0.8}$$
 of **AA** treated strain

in which 0.8 is the $\mathrm{OD}_{600\,\mathrm{nm}}$ of the untreated strain.

Growth inhibition assay of plants on plates or in the tubes. MS ($2.16 \,\mathrm{g}\,l^{-1}$ Murashige and Skoog basal medium, 8 g l⁻¹ sucrose, 8 g l⁻¹ agar) medium was used to test the growth inhibition of aspterric acid on *A. thaliana*, *S. lycopersicum* and *Z. mays* were grown under long day condition ($16/8 \,\mathrm{h}\,$ light/dark) using cool-white fluorescence bulbs as the light resource at $23\,^{\circ}\mathrm{C}$. Aspterric acid was dissolved in ethanol and added to the medium before inoculating strains or growing plants. The medium of the control treatment contained the same amount of ethanol, but without aspterric acid.

Plant growth inhibition assay by spraying. Aspterric acid was first dissolved in ethanol and then added to solvent (0.06 g l^{-1} Finale (Bayer) with 20 g l^{-1} EtOH).

The control plants were treated with solvent containing ethanol only. *A. thaliana* that are resistant to glufosinate (containing the *bar* gene) were grown under long day condition (16/8 h light/dark) using cool-white fluorescence bulbs as the light resource at 23 °C. Spraying treatments began upon seed germination and were repeated once every two days with approximately 0.4 ml aspterric acid solution each time per pot.

Structure determination of holo-*Ath***DHAD.** The gene encoding *Ath*DHAD (residues 35–608) was cloned into pET21a derivative vector pSJ2 with an eight histidine (8 \times His) tag and a TEV protease cleavage site at the N-terminus. The forward primer DHAD-F and the reverse primer DHAD-R were used for cloning. The double mutant K559A/K560A for efficient crystallization was designed using the surface entropy reduction prediction (SERp) server³⁴. Mutations were generated by PCR using the forward primer K559AK560A-F and reverse primer K559AK560A-R. All constructed plasmids were verified by DNA sequencing.

AthDHAD purified under aerobic conditions was found to contain no iron-sulfur cluster (apo form). Hence we performed [2Fe–2S] cluster reconstitution under the atmosphere of nitrogen in an anaerobic box. The protein was incubated with FeCl₃ at the ratio of 1:10 for 1 h on ice and then 10 equivalents of Na₂S per protein was added drop-wise every 30 min for 3 h. The reaction mixture was then incubated overnight. Excess FeCl₃ and Na₂S were removed using a SephadexTM G-25 Fine column (GE Healthcare)²⁶.

The reconstituted holo-AthDHAD was crystallized in an anaerobic box. The proteins (at 10 mg ml $^{-1}$) were mixed in a 1:1 ratio with the reservoir solution in a 2-µl volume and equilibrated against 50 µl reservoir solution, using the sitting-drop vapour diffusion method at 16 °C. Crystals for diffraction were observed in 0.1 M sodium acetate pH 5.0, 1.5 M ammonium sulfate after 5 days.

All crystals were flash-cooled in liquid nitrogen after cryo-protection with a solution containing 25% glycerol, 1.5 M ammonium sulfate, 0.1 M sodium acetate pH 5.0. The data were collected at 100 K at the Beam Line 19U1 in the Shanghai Synchrotron Radiation Facility (SSRF). Diffraction data of holo-AthDHAD were collected at the wavelength of 0.97774 Å. The best crystals diffracted to a resolution of 2.11 Å. The Ramachandran plot favoured (%), allowed (%) and outlier (%) are 98.05, 1.60, and 0.36, respectively. All datasets were indexed, integrated, and scaled using the HKL3000 package³⁵. The crystals belonged to space group $P4_22_12$. The statistics of the data collection are summarized in Extended Data Table 1.

The holo-*Ath*DHAD structure was solved using the molecular replacement method Phaser embedded in the CCP4i suite and the L-arabinonate dehydratase crystal structure (RCSB Protein Data Bank (PDB) ID: 5J83) as the search model. All the side chains were removed during the molecular replacement process^{36,37}. The resulting model was refined against the diffraction data using the REFMAC5 program of CCP4i³⁸. On the basis of the improved electron density, the side chains of the holo-AthDHAD protein, iron-sulfur cluster, water molecule, acetate ion, sulfate ions, and magnesium ion were manually built using the program WinCoot³⁹. The R_{work} and R_{free} values of the structure are 17.27% and 21.52%, respectively. The detailed refinement statistics are summarized in Extended Data Table 1. The geometry of the model was validated by WinCoot. Structural factor and coordinate of holo-AthDHAD have been deposited in the Protein Data Bank (PDB ID: 5ZE4). Homology modelling of AstD and docking of substrate or aspterric acid into **the active site of holo-***Ath***DHAD.** The structure of holo-*Ath***DHAD** was prepared using Schrodinger suite software under an OPLS3 force field⁴⁰. Hydrogen atoms were added to reconstituted crystal structures according to the physiological pH (7.0) with the PROPKA tool in Protein Preparation tool in Maestro to optimize the hydrogen bond network^{26,41}. Constrained energy minimizations were conducted on the full-atomic models, with heavy atom coverage to 0.5 Å. The homology model was performed in Modeller 9.18⁴², using the crystal structure of holo-AthDHAD solved in this work as a template. Sequence alignment in Modeller indicated that AstD and AthDHAD shared 56.8% sequence identity and 75.0% sequence similarity (Extended Data Fig. 6). All the highly conserved residues and motifs were properly aligned. A total of 2,000 models were generated for each target in Modeller with the fully annealed protocol. The optimal models were chosen for docking studies according to DOPE (Discrete Optimized Protein Energy) score.

All ligand structures were built in Schrodinger Maestro software²⁶. The LigPrep module in Schrodinger software was introduced for geometric optimization by using an OPLS3 force field⁴⁰. The ionization states of ligands were calculated with Epik tool using Hammett and Taft methods in conjunction with ionization and tautomerization tools⁴³. The docking of a ligand to the receptor was performed using Glide⁴⁴. We included cofactors observed in the crystal structure during the docking. As both water and SO₄²⁻ occupied the catalytic site, they were excluded before docking. Cubic boxes centred on the ligand mass centre with a radius of 8 Å for all ligands defined the docking binding regions. Flexible ligand docking was executed for all structures. Ten poses per ligand out of 20,000 were included in the post-docking energy minimization. The best scored pose for the ligand was chosen as the initial structure for further study. The molecular mechanics energies combined with the generalized Born and surface area continuum solvation

(MM/GBSA) method was introduced to evaluate the ligand binding affinity on the basis of the best-scored docking pose in Schrodinger software. Figures were prepared in PyMOL and Inkscape $^{45,46}.$ Both the native substrate $\alpha\beta$ -dihydroxyisovalerate and aspterric acid were docked into the catalytic site of AthDHAD. The cross-section electrostatic surface map shows this unique catalytic pocket has a positively charged interior and a hydrophobic entrance, which binds to negatively charged 'head' and hydrophobic 'tail' of the substrate or aspterric acid, respectively. Thus the negatively charged 'head' can lead both of the substrate and aspterric acid into the catalytic chamber. The bulky hydrophobic tricyclic moiety of aspterric acid, however, provides stronger hydrophobic interactions to the entrance and blocks the entrance of the active site owing to the hydrophobic residues at the entrance (Fig. 2d). By contrast, the smaller 'tail' of the native substrate provides fewer interactions to the entrance because the smaller size limits efficient hydrophobic contact to nearby residues. This implies that once aspterric acid binds to AthDHAD, it can prevent the substrate approaching the active site. We also introduced the MM/GBSA method, a widely used approach for relative binding energy calculation, to evaluate the relative binding affinity for both ligands⁴⁷. The MM/GBSA calculations were done in Prime⁴⁸ (Schrödinger 2015 suite). The MM/GBSA energy was calculated using following equation, $\Delta G_{\rm bind} = E_{\rm complex}$ – $E_{\rm protein}$ – Eligand. E denotes energy and includes terms such as protein—ligand van der Waals contacts, electrostatic interactions, ligand desolvation, and internal strain (ligand and protein) energies, using a VSGB2.0 implicit solvent model with the OPLS2005 force field. The solvent entropy is also included in the VSGB2.0 energy model, as it is for other generalized Born and Poison-Boltzmann continuum solvent models.

MM/GBSA calculation shows that the relative binding energy for aspterric acid and α,β -dihydroxyisovalerate is -18.6 ± 0.3 kcal mol $^{-1}$ and -13.3 ± 0.2 kcal mol $^{-1}$, respectively, which shows that the binding constant of aspterric acid to the active site is about 6,000 times greater than α,β -dihydroxyisovalerate. This further confirms that aspterric acid is a competitive inhibitor of AthDHAD. Cytotoxicity assay of aspterric acid. Cell proliferation experiments were performed in a 96-well format (five replicates per sample) using the human melanoma cell lines A375 and SK-MEL-1. Aspterric acid treatments were initiated 24 h after seeding for 72 h, and cell survival was quantified using the CellTiter-GLO assay (Promega).

Cross experiment of *A. thaliana*. To make male sterile *A. thaliana*, aspterric acid was added to a chemical hybridization agent (CHA) formulation (250 μ M aspterric acid, 2% ethanol, 0.1% Tween-80, 1% corn oil in water), which has less inhibition effect on the growth of *A. thaliana*. Flowers of the aspterric acid-treated Col-0 were selected as the female parent. The non-treated *A. thaliana* containing a glufosinate resistance gene were used as the male parent to donate pollen. Two-week-old F1 progeny resulting from the cross were treated by Finale (11.3% glufosinate-ammonium) at a 1:2,000 dilution. The results are summarized in Extended Data Fig. 9.

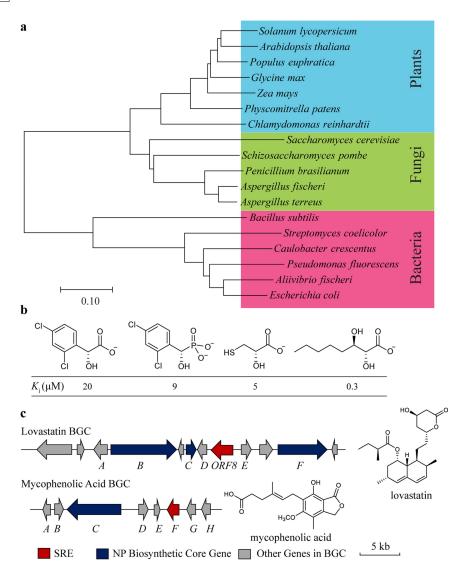
Construction of the transgenic plants. The coding sequence of AstD was codon optimized for *A. thaliana*. A chloroplast-localization signal (CLS) of 35 amino acid residues derived from the N-terminal of *A. thaliana* DHAD (MQATIFSPRATLFPCKPLLPSHNVNSRRPSIISCS) was fused to the N terminus of the codon-optimized AstD. A 3 × Flag-tag was inserted between the CLS and the codon-optimized AstD (Supplementary Table 6). The gene block containing CLS, the Flag-tag and *astD* was synthesized and then cloned into pEG202 vector using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific). The original CaMV 35S promoter of pEG202 was substituted by the ubiquitin-10 promoter to drive the expression of AstD. The construct was electro-transformed into *Agrobacterium tumefaciens* strain Agl0 and then transformed into *A. thaliana* using the standard floral dip method⁴⁹. The *A. thaliana* Col-0 ecotype was transformed. Positive transgenic plants were selected using the glufosinate resistance marker, and were tested for survival in the presence of aspterric acid.

Protein expression verification with western blot. Approximately 0.5 g of leaf tissue of transgenic A. thaliana was ground in liquid nitrogen. Proteins were homogenized in $2 \times SDS$ buffer and then centrifuged at 21,000g for 5 min to remove undissolved debris. The supernatant containing resolved proteins were loaded onto a 4-12% Bis–Tris gel, and separated using MOPS running buffer. Transfer was conducted using an iBlot2 dry transfer device and a PVDF membrane. The total proteins were stained with Ponceau to demonstrate equal loading. Western blotting was performed using Sigma monoclonal anti-Flag M2-Peroxidase antibody, with detection using the Amersham ECL Prime detection reagent.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data that support the findings of this study are available within the paper and its Supplementary Information, or are available from the corresponding authors upon reasonable request. The structural factor and coordinate of holo-*Ath*DHAD have been deposited in the Protein Data Bank under the ID 5ZE4.

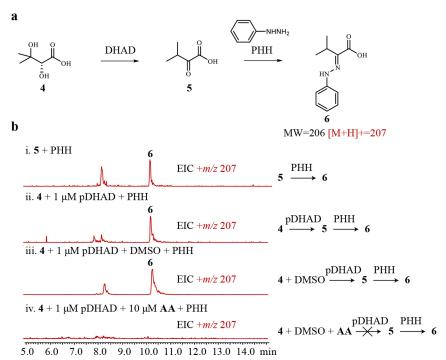
- Liu, N. et al. Identification and heterologous production of a benzoyl-primed tricarboxylic acid polyketide intermediate from the zaragozic acid A biosynthetic pathway. Org. Lett. 19, 3560–3563 (2017).
- 32. Fang, F. et al. A vector set for systematic metabolic engineering in Saccharomyces cerevisiae. Yeast 28, 123–136 (2011).
- Cool, L. G. ent-Daucane and acorane sesquiterpenes from × Cupressocyparis leylandii foliage. Phytochemistry 58, 969–972 (2001).
- Góldschmidt, L., Cóoper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: a web server for the design of crystallizable protein variants. *Protein Sci.* 16, 1569–1576 (2007).
- Otwinowski, Z., Minor, W. & Jr, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326 (1997).
- McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. D 67, 235–242 (2011).
- Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D 67, 355–367 (2011).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).
- Harder, E. et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J. Chem. Theory Comput. 12, 281–296 (2016).
- Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. J. Chem. Theory Comput. 7, 2284–2295 (2011).
- Eswar, N. et al. Comparative protein structure modeling using MODELLER. Curr. Protoc. Bioinformatics 47, 5.6.1–5.6.32 (2006).
- Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* 24, 591–604 (2010).
- Friesner, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749 (2004).
- Yuan, S., Chan, H. C. S., Filipek, S. & Vogel, H. PyMOL and Inkscape bridge the data and the data visualization. Structure 24, 2041–2042 (2016).
- Yuan, S., Chan, H. C. S. & Hu, Z. Using PyMOL as a platform for computational drug design. Wiley Interdiscip. Rev. Comput. Mol. Sci. 7, e1298 (2017).
- 47. Genheden, S. & Rýde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discov. 10, 449–461 (2015).
- Sirin, S. et al. A computational approach to enzyme design: predicting ω-aminotransferase catalytic activity using docking and MM-GBSA scoring J. Chem. Inf. Model. 54, 2334–2346 (2014).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacteriummediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743 (1998).



Extended Data Fig. 1 | The rationale of resistance-gene-directed discovery of a natural herbicide with a new mode of action.

a, Phylogenetic tree of DHAD among bacteria, fungi and plants. The evolutionary history was inferred by using the neighbour-joining method (MEGA7). Scale-bar units represent the number of amino acid substitutions per site. b, Representatives of small molecules that inhibit DHAD in vitro, but fail to inhibit plant growth. c, Examples of co-localization of biosynthetic gene clusters (BGCs) and targets.

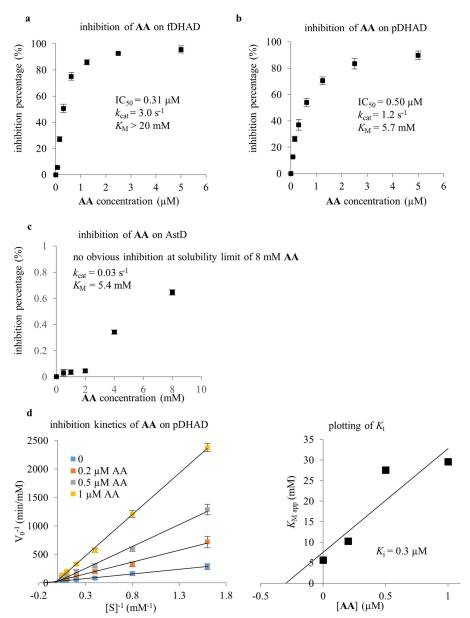
The biosynthetic core genes are shown in blue and the self-resistance enzymes (SREs) are shown in red. The blockbuster cholesterol-lowering lovastatin drug targets HMG-CoA reductase (HMGR) in eukaryotes. In the fungus *A. terreus* that produces lovastatin, a second copy of HMGR encoded by ORF8 is present in the gene cluster (top). The BGC of the immunosuppressant mycophenolic acid from *Penicillium* sp. contains a second copy of inosine monophosphate dehydrogenase (IMPDH), which represents the SRE to this cluster (bottom).



Extended Data Fig. 2 \mid Biochemical assays of DHAD functions.

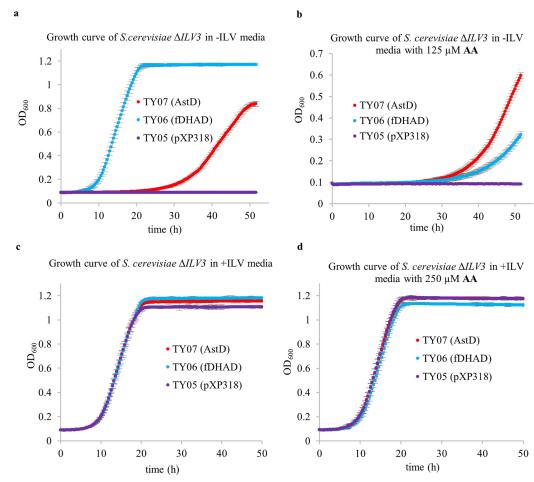
a, Assaying DHAD activities in the conversion of the dihydroxyacid 4 into the α -ketoacid 5. Formation of 5 can be detected with HPLC by chemical derivatization using phenylhydrazine (PHH) to yield 6. b, LC–MS traces of the biochemical assays of AthDHAD (plant DHAD, pDHAD). EIC of positive ion mass of $[M+H]^+=207$ is shown in red. Panels i–iv in b: i,

the derivatization reaction was validated by using the authentic 5; ii, the bioactivity of AthDHAD in converting 4 into 5 was validated; iii, addition of DMSO to AthDHAD enzymatic reaction mixture has no effect; and iv, addition of $10~\mu M$ aspterric acid to the reaction mixture abolished AthDHAD activity. The experiments were repeated independently three times with similar results.



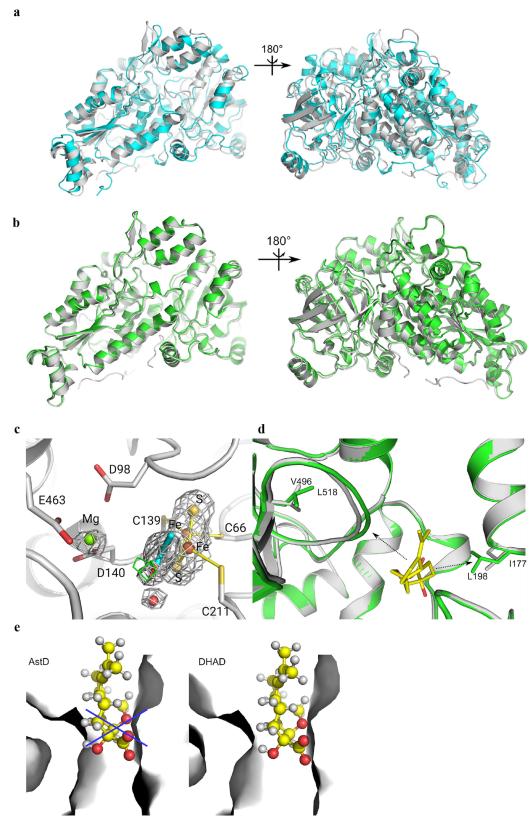
Extended Data Fig. 3 | Inhibition assay of different DHADs using aspterric acid. a–c, Three DHAD enzymes were assayed, including AthDHAD (plant DHAD, pDHAD), AteDHAD (fungal housekeeping DHAD from A. terreus, fDHAD) and AstD (DHAD homologue within ast cluster). IC $_{50}$ and K_i values of aspterric acid were measured on the basis of inhibition percentage at different aspterric acid concentrations. Data are mean \pm s.d. from three biologically independent experiments. a, Plot of the inhibition percentage of 0.5 μ M AteDHAD as a function

of aspterric acid concentration. **b**, Plot of the inhibition percentage of 0.5 μ M AthDHAD as a function of aspterric acid concentration. **c**, Plot of the inhibition percentage of 0.5 μ M AstD as a function of aspterric acid concentration. **d**, Analysis of inhibitory kinetics of aspterric acid on AthDHAD using the Lineweaver–Burk method at different concentrations of aspterric acid (left). Linear fitting of the apparent Michaelis constant ($K_{\rm M,app}$) as a function of aspterric acid concentration yields the $K_{\rm i}$ of aspterric acid on AthDHAD (right).



Extended Data Fig. 4 | Growth curve of *S. cerevisiae* $\Delta ILV3$ expressing AstD and AteDHAD. a–d, The genome copy of DHAD encoded by ILV3 was first deleted from *S. cerevisiae* strain DHY $\Delta URA3$ to give UB02. UB02 was then either chemically complemented by growth on ILV (leucine, isoleucine and valine)-containing medium or genetically by expressing of AteDHAD (fungal housekeeping DHAD from A. terreus, fDHAD) or AstD episomally (TY06 or TY07, respectively). The empty

vector pXP318 was also transformed into UB02 to generate a control strain TY05. Cell growth (optical density) under different conditions was plotted as a function of time. Data are mean \pm s.d. from three biologically independent experiments. a, Growth curve in ILV dropout medium with no aspterric acid. b, Growth curve in ILV dropout medium with 125 μM aspterric acid. c, Growth curve in ILV supplemented medium. d, Growth curve in ILV supplemented medium with 250 μM aspterric acid.



Extended Data Fig. 5 \mid See next page for caption.

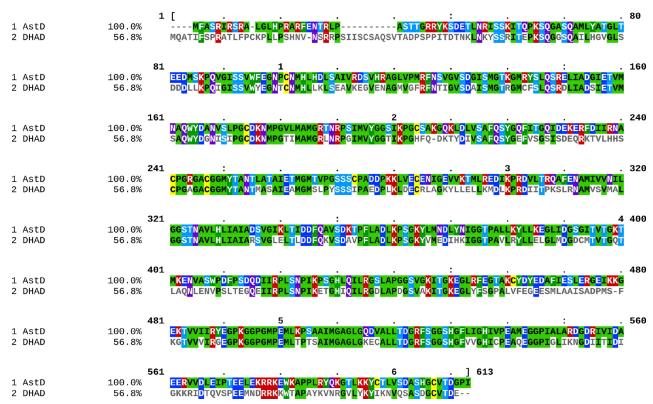


Extended Data Fig. 5 | X-ray structure of holo-AthDHAD and homology model of AstD. a, Superimpositions of the monomer of holo-AthDHAD (PDB: 5ZE4, 2.11 Å) and RlArDHT (PDB: 5J84). The holo structure containing the 2Fe–2S cofactor and Mg²⁺ ion in the active site. The structure of holo-AthDHAD is in white; the crystal structure of RlArDHT is in cyan. b, Superimpositions of holo-AthDHAD and homology-modelled AstD. The structure of AstD was constructed by homology modelling on the basis of the structure of holo-AthDHAD. The structure of holo-AthDHAD is in white; the crystal structure of AstD is in green. c, The electron density map of cofactors in the holo structure of AthDHAD. White mesh indicates the $2F_0 - F_c$ map at the 1.2σ level;

green mesh indicates the $F_{\rm o}-F_{\rm c}$ positive map at the 3.2 σ level; cyan sticks represent the acetic acid molecule. **d**, Comparison of the active sites in the crystal structure of AthDHAD and the modelled structure of AstD. The cartoon represents superimposed binding sites of AthDHAD (white) and AstD (green). The shift of a loop in AstD, where L518 (corresponding to V496 in AthDHAD) is located, coupled with a larger L198 residue (corresponding to 1177 in AthDHAD) leads to a smaller hydrophobic pocket in AstD than in AthDHAD. **e**, The surface of binding sites of AstD (left) and AthDHAD (right). The smaller hydrophobic channel in the modelled AstD cannot accommodate the aspterric acid molecule (yellow ball and stick model).



Identities normalised by aligned length. Colored by: identity + property



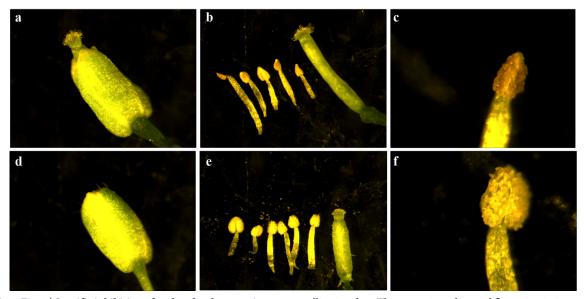
Extended Data Fig. 6 | **Sequence alignment between** *Ath***DHAD and AstD.** The sequence identity between *Ath*DHAD and AstD is 56.8%, whereas the similarity between them is 75.0%. Residues were coloured according to their property and similarity.



Solvent

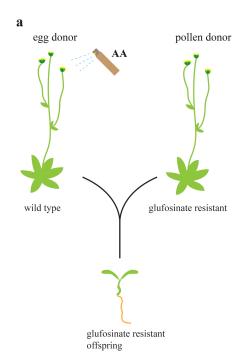
 $250 \mu M AA$ in solvent

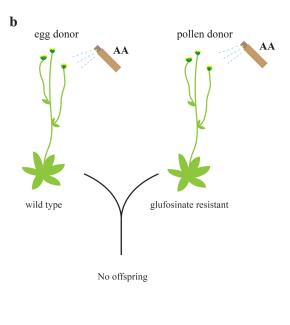
Extended Data Fig. 7 | Spray assay of aspterric acid on A. thaliana. Glufosinate-resistant A. thaliana was treated with (right) or without (left) aspterric acid in the solvent, which is a commercial glufosinate-based herbicide marketed as Finale. To improve the wetting and penetration, aspterric acid was first dissolved in ethanol and then added to the solvent (0.06 g l $^{-1}$ Finale (Bayer) with 20 g l $^{-1}$ ethanol) to make 250 μ M aspterric acid spraying solution. The control plants were treated with solvent containing ethanol only. Spraying treatments began upon seed germination, and were repeated once every two days with approximately 0.4 ml aspterric acid solution per time per pot for four weeks. The picture shown is taken after one month of treatment. The application rate of aspterric acid is approximately 1.6 lb per acre, which is comparable to the commonly used herbicide glyphosate (0.75–1.5 lb per acre). The experiments were repeated independently three times with similar results.



Extended Data Fig. 8 | Specific inhibition of anther development in A. thaliana. a-f, Comparison of flower organs between the aspterric acid-treated $(\mathbf{a}-\mathbf{c})$ and non-treated $(\mathbf{d}-\mathbf{f})$ Arabidopsis. a, d, The aspterric acid-treated flower shows abnormal pistil elongation owing to the lack of

pollination. **b**, **e**, The aspterric acid treated flower is missing one stamen. **c**, **f**, The aspterric acid treated anther is depleted of healthy and mature pollen. The experiments were performed twice with similar results.





female parent	male parent	offspring obtained	inherit resistance
AA treated wild type	un-treated	Yes	Yes
	Glufosinate resistant		
	plant		
AA treated wild type	AA treated	No	N/A
	glufosinate resistant		
	plant		

Extended Data Fig. 9 | Schematic of results from the cross experiment. a, Wild-type *A. thaliana* treated with 250 μ M aspterric acid was pollinated with pollen from the un-treated plant that carries the glufosinateresistance gene. Offspring was obtained, and inherited the glufosinate

resistance from the pollen donor. b, As in a, except that the pollen donor was also treated with 250 μM aspterric acid. No offspring was obtained from this cross. Similar results were obtained after treatment with 100 μM aspterric acid.

Extended Data Table 1 | Data collection and refinement statistics (molecular replacement)

	holo- <i>Ath</i> DHAD
Data collection	
Space group	$P4_{2}2_{1}2$
Cell dimensions	
a,b,c (Å)	135.5, 135.5, 66.0
α, β, γ (°)	90, 90, 90
Resolution (Å)	47.89-2.11 (2.15-2.11) *
R_{sym} or R_{merge}	0.189 (1.240)
$I/\sigma I$	17.86 (2.33)
Completeness (%)	100 (100)
Redundancy	25.1 (23.1)
Refinement	
Resolution (Å)	30.00-2.11
No. reflections	33076 (1709)
$R_{ m work}$ / $R_{ m free}$	0.1727 / 0.2152
No. atoms	
Protein	4224
Ligand/ion	24
Water	118
B-factors	
Protein	26.60
Ligand/ion	46.53
Water	26.22
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.191

^{*}Values in parentheses are for the highest-resolution shell.



Prespliceosome structure provides insights into spliceosome assembly and regulation

Clemens Plaschka $^{1,2,3*},$ Pei–Chun Lin $^{1,3*},$ Clément Charenton 1 & Kiyoshi Nagai 1*

The spliceosome catalyses the excision of introns from pre-mRNA in two steps, branching and exon ligation, and is assembled from five small nuclear ribonucleoprotein particles (snRNPs; U1, U2, U4, U5, U6) and numerous non-snRNP factors¹. For branching, the intron 5' splice site and the branch point sequence are selected and brought by the U1 and U2 snRNPs into the prespliceosome¹, which is a focal point for regulation by alternative splicing factors². The U4/U6.U5 tri-snRNP subsequently joins the prespliceosome to form the complete pre-catalytic spliceosome. Recent studies have revealed the structural basis of the branching and exon-ligation reactions³, however, the structural basis of the early events in spliceosome assembly remains poorly understood⁴. Here we report the cryo-electron microscopy structure of the yeast Saccharomyces cerevisiae prespliceosome at near-atomic resolution. The structure reveals an induced stabilization of the 5' splice site in the U1 snRNP, and provides structural insights into the functions of the human alternative splicing factors LUC7-like (yeast Luc7) and TIA-1 (yeast Nam8), both of which have been linked to human disease^{5,6}. In the prespliceosome, the U1 snRNP associates with the U2 snRNP through a stable contact with the U2 3' domain and a transient yeast-specific contact with the U2 SF3b-containing 5' region, leaving its tri-snRNP-binding interface fully exposed. The results suggest mechanisms for 5' splice site transfer to the U6 ACAGAGA region within the assembled spliceosome and for its subsequent conversion to the activation-competent B-complex spliceosome^{7,8}. Taken together, the data provide a working model to investigate the early steps of spliceosome assembly.

To gain structural insights into early spliceosome assembly, we prepared the yeast prespliceosome A-complex on the *UBC4* pre-mRNA

that carries a mutation in the pre-mRNA branch point sequence, which was previously used to stall the A-complex⁹ (UACUAAC to UACAAAC, in which **A** is the branch point adenosine and \underline{A} is the mutated nucleotide) (Extended Data Fig. 1a, b). The purified A-complex contained stoichiometric amounts of the U1 and U2 snRNP proteins (Extended Data Fig. 1b), and was used to determine cryo-electron microscopy (cryo-EM) densities of the A-complex at 4.0 Å (U1 snRNP, map A2) and 4.9-10.4 Å (U2 snRNP, maps A1 and A3) resolution, respectively (Extended Data Figs. 1c-e, 2). From these densities we built a near-complete atomic model of the A-complex (Fig. 1, Supplementary Videos 1, 2, Supplementary Data, Extended Data Fig. 1f), comprising 34 proteins, U1 and U2 snRNAs, and 34 nucleotides of pre-mRNA. The final model lacks the mobile cap-binding complex, Prp5 or the U1 subunit Prp40 (Extended Data Fig. 1b, d, e; Extended Data Table 1). The elongated U1 and U2 snRNPs bind the pre-mRNA 5' splice site (5'SS) and branch-point sequences, respectively, and associate in a parallel manner to form the A-complex (Fig. 2a). The U1 snRNP structure contains all the essential regions of the U1 snRNA and 16 proteins (Fig. 1). The U1 snRNP 'core' is highly similar to its human counterpart¹⁰ (Extended Data Figs. 3, 4), comprising the seven-membered Sm ring and orthologues of the human U1 snRNP proteins (Snp1, human U1-70k; Mud1, human U1A; Yhc1, human U1C), and is bound to the peripheral yeast U1 proteins Luc7, Nam8, Prp39, Prp42, Snu56 and Snu7111 (Extended Data Figs. 3, 4). The U2 snRNP has a bipartite structure as observed in B-complex⁸, comprising the SF3b subcomplex ('5' region') and the U2 3' domain and SF3a subcomplex ('3' region') that are organized around the 5' and 3' regions of the U2 snRNA, respectively (Figs. 1, 2a, Extended Data Fig. 5). At the current resolution, the conformation of the U2 5' region appears unchanged from the

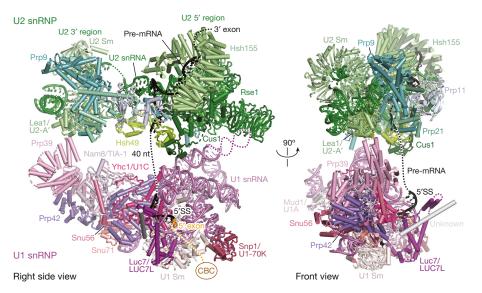


Fig. 1 | Prespliceosome A-complex structure. Two orthogonal views of the yeast A-complex structure. Subunits are coloured according to snRNP identity (U1, shades of purple, U2, shades of green), and the pre-mRNA intron (black) and its 5' exon (orange) are highlighted. The orthologous human protein name is shown after the solidus. The location of the cap-binding complex (CBC) is indicated by a brown oval (see Extended Data Fig. 1e).

¹MRC Laboratory of Molecular Biology, Cambridge, UK. ²Present address: Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria. ³These authors contributed equally: Clemens Plaschka, Pei-Chun Lin. *e-mail: clemens.plaschka@imp.ac.at; pclin@mrc-lmb.cam.ac.uk; kn@mrc-lmb.cam.ac.uk

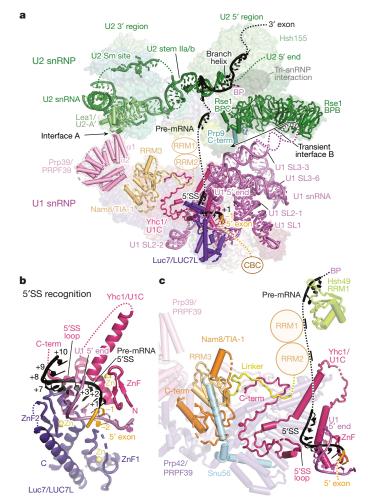


Fig. 2 | 5'SS recognition and implications for alternative splicing. a, The A-complex U1-U2 snRNP interfaces (A and B) and the RNA network are shown as cartoons, and are superimposed on the transparent surfaces of the prespliceosome proteins. The U2 subunit Hsh155 surface (grey oval), which interacts with the tri-snRNP in the B-complex, is freely accessible in the A-complex. The U1 snRNP proteins Nam8 (orange, human TIA-1), Luc7 (purple, human LUC7L), Prp39 (magenta, human PRPF39) and Yhc1 (dark magenta, human U1C) and the U2 snRNP proteins Lea1 (light green, human U2-A'), Rse1 (dark green, human SF3B3), and Prp9 (teal, human SFA3) are shown as ribbons. BP, branch point. b, The pre-mRNA 5'SS is recognized by the U1 snRNA 5' end, and is stabilized by Luc7 and Yhc1. Notably, the Yhc1 ZnF and Luc7 ZnF2 domains are arranged with pseudo-C2 symmetry around the U1-5'SS helix. c, Nam8 binds the U1 snRNP through its linker (yellow), RNA recognition motif 3 (RRM3, light orange) and C-terminal regions (orange), whereas its RRM1 and RRM2 domains are mobile and project towards the intron to bind uridine-rich sequences downstream of the pre-mRNA 5'SS (dashed black line), as with its human counterpart TIA-1¹⁸. Nam8 contacts the Yhc1 (human U1C) C terminus, and human TIA-1 biochemically also interacts with human U1C¹⁸. Snu56 (blue), Prp39 (magenta), Prp42 (violet), and Hsh49 (light green) are shown as transparent ribbon models and other protein and U1 snRNA elements were removed for clarity.

B-complex⁸, in which the pre-mRNA branch-point sequence is base-paired with the U2 snRNA and the branch point adenosine is bulged out and accommodated in a pocket formed by the U2 SF3b subunits Hsh155 and Rds3. After we completed the A-complex structure, the cryo-EM structure of the free yeast U1 snRNP was reported¹². This model is in good agreement with the U1 snRNP in our A-complex structure, but there are important differences¹².

The first ten nucleotides of U1 snRNA are disordered in the free U1 snRNP¹², but become ordered in our A-complex structure by pairing with the pre-mRNA 5'SS (Fig. 2a, b). Additional density appeared

adjacent to the U1-5'SS helix, into which we could build a newly ordered Yhc1 peptide (human U1C) that contacts the 5'SS phosphate backbone (+5 and +6 positions, the 'Yhc1-5'SS loop') and a nearcomplete model of Luc7 (in the previous study Luc7 was attributed to what is now assigned as Snu7112) (Extended Data Figs. 3a, c, 4a). Although Luc7 is disordered in the free U1 snRNP, it associates stably with the U1-5'SS helix in the A-complex (Extended Data Fig. 4a), suggesting a mechanism for the selection of weak 5'SS sequences¹³. In our structure Luc7 is anchored by its N-terminal α -helix 1 to the Sm ring subunit SmE, and its C₃H-type zinc finger 1 (ZnF1) domain binds where the 5' exon emerges from the U1–5'SS helix, in excellent agreement with RNA-protein crosslinks¹³ (Fig. 2b). The adjacent Luc7 C_2H_2 -type ZnF2 contacts the U1-5'SS helix minor groove and the U1 snRNA phosphate backbone (nucleotides U5–C8). This interaction mirrors that between the Yhc1 ZnF domain and the 5'SS nucleotides +1 to +4 downstream of the 5'SS junction¹⁰ (Fig. 2b). Thus, Yhc1 and Luc7 make no base-specific interactions with the U1-5'SS helix, and instead cradle the U1-5'SS helix phosphate backbone to stabilize 5'SS binding. Consistent with the structure, weakening of any of these interactions can impair splicing and bypass the requirement for Prp28 helicase activity^{13–16}.

The A-complex structure reveals structural insights into the functions of the human alternative splicing factors LUC7-like (LUC7L, yeast Luc7) and TIA-1 (yeast Nam8) (Extended Data Fig. 4c, d). Luc7 and its human homologues LUC7L1-3 are highly conserved, suggesting that the LUC7L N-terminal α -helix also anchors it to the SmE protein and that the invariant ZnF2 helix $\alpha 8$ similarly stabilizes the U1-5'SS helix to promote the inclusion of weak alternative splice sites¹³ (Fig. 2b, Extended Data Figs. 3c, 6a). The yeast U1 snRNP subunit Nam8 and its human homologue TIA-1 contain three RNA recognition motif (RRM) domains and a C-terminal Gln-rich extension (Extended Data Fig. 6b). Human TIA-1 binds to uridine-rich sequences downstream of the 5'SS predominantly through the RRM2 domain 17,18 to allow the use of weak 5'SSs. The Nam8 RRM2 shows high sequence similarity to the TIA-1 RRM2, including the nearly identical RNP1 and RNP2 motifs, indicating that Nam8 also binds uridine-rich sequences through its RRM2 also (Extended Data Fig. 6b). In the A-complex structure the Nam8 RRM3 and its C-terminal region bind in a cavity of the Prp39–Prp42 heterodimer and contact the Yhc1 C-terminal region near the U1–5/SS helix (Fig. 2c). From this location, Nam8 could project its mobile RRM2 domain to bind uridine-rich intron sequences downstream of the 5'SS, consistent with crosslinking experiments¹⁷, and thereby promote meiotic pre-mRNA splicing¹⁹ (Fig. 2a, c).

In the A-complex, the U1 snRNP binds to the U2 snRNP through two interfaces, A and B (Fig. 2a). In interface A, the N-terminal helices α 1–2 of the U1 protein Prp39 stably bind the U2 3' domain subunit Lea1 (human U2A') (Fig. 2a, Extended Data Fig. 5). The Prp39–Prp42 heterodimer binds Yhc1 to anchor the U2 snRNP 3' domain to the U1 snRNP. Similar interactions were observed biochemically between the human alternative-splicing factor PRPF39 homodimer and U1C¹² (yeast Yhc1), suggesting that PRPF39 may contact the human U2 3' domain in a similar manner, although it is not an obligate component of the human A-complex²⁰ (Fig. 2a). Different, non-overlapping Lea1 surfaces are used to interact with the NTC protein Syf1 in the yeast C- and C*/P-complex conformations of the spliceosome²¹ (Extended Data Fig. 5c), suggesting that Lea1 aids in the repositioning of the U2 3' domain in multiple stages of splicing. Interface B is transient and found only in a subset of cryo-EM images (Extended Data Figs. 2a, 5a, b). It involves weak interactions between the yeast-specific U1 snRNA stem loop 3–3 and the U2 SF3b Rse1 subunit β -propellers B and C (BPB and BPC) and the C terminus of U2 SF3a Prp9. The pre-mRNA 5'SS and branch point branching reactants are positioned approximately 150 Å apart in the A-complex, with 40 nucleotides of the UBC4 intron looped out in between (Fig. 2a, Extended Data Fig. 1e, f). The small interfaces between the U1 and U2 snRNPs orient the snRNPs relative to each other, and this may facilitate 5'SS transfer in the assembled spliceosome and the subsequent dissociation of the U1 snRNP, consistent with the

Fig. 3 | Spliceosome assembly and 5'SS transfer. a, One of the two alternative pre-B-complex models, suggesting that the U2 snRNP orients the U1 snRNP to deliver the pre-mRNA 5'SS to the U6 ACAGAGA stem. The model was obtained by superposing the yeast A- (from this study) and B-complex structures (RCSB Protein Data Bank code (PDB ID) 5NRL) and by modifying the locations of Brr2, U4 Sm ring, Sad1, and Prp28 to resemble a human-like pre-B-complex conformation on the basis of the biochemical data and the human U4/U6.U5 tri-snRNP structure (PDB ID 3JCR) (see 'Structural modelling' in Methods). Colouring as in Fig. 1

and a previously published work⁸. **b**, The pre-B-complex RNA network and the Prp28 helicase are shown as cartoons and are superimposed on transparent surfaces of the spliceosome proteins. Prp28 is positioned at the Prp8 N-terminal domain as in human tri-snRNP²⁵ and may clamp onto the pre-mRNA near the U1-5'SS helix to destabilize it and transfer the 5'SS from U1 snRNA to the U6 snRNA ACAGAGA stem (red arrow), which are separated by approximately 20 Å in the pre-B model. The positions of proteins marked with asterisks are based on the human tri-snRNP structure (PDB ID 3JCR).

structural and biochemical data^{7,8}. Although the precise U1–U2 snRNP interfaces may differ in the human A-complex, a key function of U1–U2 (alternative) splicing factors could be to ensure that U1 and the U1–5'SS helix are oriented correctly relative to the U2 snRNP.

Before A-complex formation, the yeast Msl5-Mud2 heterodimer recognizes the branch point sequence through Msl5 and binds the U1 snRNP subunit Prp40 (human PRPF40) in the E complex, looping out the intron between the 5'SS and branch point sequences²² (Extended Data Fig. 4e). Although Prp40 was not identified in the free U1 snRNP¹² or in our A-complex structure, Prp40 crosslinks to Luc7 and Snu7112 and unassigned cryo-EM density in the A-complex may indicate its peripheral location near Luc7 (Extended Data Figs. 1e, 4a, e). Msl5–Mud2 may then be destabilized by the Sub2 helicase, allowing the Prp5 helicase to remodel U2 snRNA for the stable association of the U2 snRNP with the branch point sequence in the A-complex⁹. Prp5 was shown to physically interact with the U2 SF3b subunit Hsh155 HEAT repeats 1-6 and 9-12²³ and with U2 snRNA at and surrounding the branch point-interacting stem loop⁹. Thus, after Prp5 activity, Prp5 needs to dissociate to fully expose the Hsh155 HEAT repeats 11-13 together with the U2 snRNA 5' end in the A-complex, to allow for the subsequent U4/U6.U5 tri-snRNP association to assemble the spliceosome $^{7-9}$ (Fig. 2a).

The A-complex structure also provides new insights into formation of the fully assembled pre-B-complex spliceosome, which requires integration of the tri-snRNP with the A-complex. The subsequent Prp28 helicase-mediated transfer of the 5'SS from U1 to U6 snRNA and destabilization of the U1 snRNP produces the B-complex spliceosome²⁴. We first modelled a fully assembled yeast spliceosome, by superimposing the U2 snRNP SF3b-containing domains of the yeast A-complex (from this study) and the yeast B-complex structure⁸. As in the B-complex structure8, the U2 snRNP would associate with tri-snRNP via the U2/U6 helix II and Prp3 (Extended Data Fig. 7). The modelling shows that the U1 snRNP would clash with large parts of the Brr2-containing 'helicase' domain ('U1-B-complex'; Extended Data Figs. 7b, 8b), which may be relieved owing to their known flexibilities⁸ (Extended Data Fig. 5a). However the known binding site for Prp28 at the U5 Prp8 N-terminal domain (Prp8^N) observed in human tri-snRNP²⁵ would be sterically occluded by the pre-bound B-complex proteins ^{7,8,26}. We therefore considered an alternative model for the assembled yeast 'pre-B-complex' spliceosome, by combining the available data from yeast and human systems 8,25,27,28 (Fig. 3a, Extended Data Figs. 7a, 8a). First, the isolated human²⁵ and yeast tri-snRNP^{26,29}

structures differ in their protein composition and conformation, indicating that different complexes accumulate at steady-state. In the human tri-snRNP structure²⁵ the BRR2 helicase is held near SNU114 by the SAD1 protein and PRP28 is bound to the PRP8 N-terminal domain. In the yeast tri-snRNP 26,29 and the yeast and human B-complex structures^{7,8} Brr2 is repositioned and loaded onto its U4 snRNA substrate and the B-complex proteins replace Prp28 at the Prp8^N domain, ready for spliceosome activation. Second, in humans, an ATPase-deficient PRP28 helicase stalls spliceosome assembly at the pre-B-complex stage, before disruption of the U1-5'SS interaction²⁸ and this complex comprises the U1 and U2 snRNPs, a loosely associated tri-snRNP, and SAD1²⁸. Third, in yeast, Sad1 is essential for splicing and is very transiently associated with the tri-snRNP²⁷. Given the high conservation of the major spliceosome components in yeast and humans, the yeast spliceosome may likewise assemble with a human-like tri-snRNP that contains Prp28, Sad1 and a repositioned Brr2 helicase^{25,28}. On the basis of these assumptions, we modelled a yeast pre-B-complex spliceosome that comprises all five snRNPs with a combined molecular mass of approximately 3.1 megadalton and with only minor clashes (Fig. 3a, Extended Data Fig. 7a, b). Notably, this model indicates that the U2 snRNP positions the U1 snRNP to deliver the U1-5'SS helix to the exposed U6 ACAGAGA stem in tri-snRNP, only approximately 20 Å away from where Prp28 is likely to mediate 5'SS transfer, consistent with protein-RNA crosslinks³⁰ (Fig. 3b). This suggests that the subsequent repositioning of the Brr2 helicase onto the U4 snRNA, observed in the B complex structure^{7,8}, would coincide with the release of the U1 snRNP owing to a steric clash, rendering Brr2 competent for spliceosome activation only after successful 5'SS transfer (Extended Data Figs. 7b, 8a). The model thus indicates a new molecular checkpoint to couple 5'SS transfer with U1 snRNP release and formation of the B-complex (Extended Data Figs. 7b, 8a).

In summary, the prespliceosome structure reveals how the U1 and U2 snRNPs recognize the two reactants of the branching reaction and associate together with the tri-snRNP into the fully assembled spliceosome. The results further suggest how the human alternative-splicing factors LUC7L and TIA-1 may influence splice-site selection.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0323-8.



Received: 11 January; Accepted: 16 May 2018; Published online 11 July 2018.

- Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb. 1. Perspect. Biol. 3, a003707 (2011).
- Papasaikas, P. & Valcárcel, J. The spliceosome: the ultimate RNA chaperone and sculptor. Trends Biochem. Sci. 41, 33-45 (2016).
- Wilkinson, M. E., Lin, P. C., Plaschka, C. & Nagai, K. Cryo-EM studies of pre-mRNA splicing: from sample preparation to model visualization. *Annu. Rev. Biophys.* **47**, 175–199 (2018).
- Behzadnia, N. et al. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. 26, 1737–1748
- Gao, G. & Dudley, S. C. Jr. RBM25/LUC7L3 function in cardiac sodium channel splicing regulation of human heart failure. Trends Cardiovasc. Med. 23, 5-8
- Hackman, P. et al. Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. Ann. Neurol. 73, 500-509 (2013).
- Bertram, K. et al. Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell 170, 701-713.e711 (2017).
- Plaschka, C., Lin, P. C. & Nagai, K. Structure of a pre-catalytic spliceosome.
- Nature 546, 617–621 (2017). Liang, W. W. & Cheng, S. C. A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. Genes Dev. 29, 81-93 (2015).
- 10. Kondo, Y., Oubridge, C., van Roon, A. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. eLife 4, e04986 (2015).
- 11. Neubauer, G. et al. Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc. Natl Acad. Sci. USA 94, 385-390 (1997)
- 12. Li, X. et al. CryoEM structure of Saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nat. Commun. 8, 1035 (2017).
- 13. Puig, O., Bragado-Nilsson, E., Koski, T. & Séraphin, B. The U1 snRNP-associated factor Luc7p affects 5' splice site selection in yeast and human. Nucleic Acids Res. 35, 5874-5885 (2007).
- 14. Agarwal, R., Schwer, B. & Shuman, S. Structure-function analysis and genetic interactions of the Luc7 subunit of the Saccharomyces cerevisiae U1 snRNP. RNA 22, 1302-1310 (2016).
- 15. Schwer, B. & Shuman, S. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1 Tgs1, U1 snRNA, SmD3 and Prp28. *Nucleic Acids Res.* **42**, 4697–4711 (2014).
- 16. Chen, J. Y. et al. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol. Cell 7, 227-232 (2001).
- 17. Puig, O., Gottschalk, A., Fabrizio, P. & Séraphin, B. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. Genes Dev. 13, 569-580 (1999).
- 18. Förch, P., Puig, O., Martínez, C., Séraphin, B. & Valcárcel, J. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J. 21, 6882-6892 (2002).
- 19. Spingola, M. & Ares, M. Jr. A yeast intronic splicing enhancer and Nam8p are required for Mer1p-activated splicing. Mol. Cell 6, 329-338 (2000).
- Agafonov, D. E. et al. ATP-\S stalls splicing after B complex formation but prior to spliceosome activation. *RNA* **22**, 1329–1337 (2016).

- 21. Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* **24**, 791–799 (2017).
- 22. Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89, 403-412 (1997).
- 23. Tang, Q. et al. SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev. 30, 2710-2723 (2016).
- Staley, J. P. & Guthrie, C. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol. Cell 3, 55-64 (1999).
- Agafonov, D. E. et al. Molecular architecture of the human U4/U6.U5 tri-snRNP. Science 351, 1416-1420 (2016).
- Nguyen, T. H. D. et al. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. Nature **530**, 298–302 (2016).
- Huang, Y. H., Chung, C. S., Kao, D. I., Kao, T. C. & Cheng, S. C. Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. Mol. Cell. Biol. 34, 210-220 (2014).
- Boesler, C. et al. A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. Nat. Commun. 7, 11997 (2016).
- Wan, R. et al. The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**, 466–475 (2016).

 30. Ismaïli, N., Sha, M., Gustafson, E. H. & Konarska, M. M. The 100-kda U5 snRNP
- protein (hPrp28p) contacts the 5' splice site through its ATPase site. RNA 7, 182-193 (2001).

Acknowledgements We thank C. Savva, S. Chen, G. Cannone, G. McMullan, J. Grimmett and T. Darling for maintaining electron microscopy and computing facilities; the mass spectrometry facility for protein identification; A. Murzin for discussions; E. Hesketh and R. Thompson for assistance with cryo-EM data collection of dataset three and S.-C. Cheng, A. Newman, L. Strittmatter, M. E. Wilkinson for critical reading of the manuscript. We thank J. Löwe, V. Ramakrishnan, D. Barford and R. Henderson for their continuing support. The project was supported by the Medical Research Council (MC_U105184330) and European Research Council Advanced Grant (693087-SPLICE3D). C.P. was supported by an EMBO Long-Term Fellowship (984-2015).

Author contributions C.P. and P.-C.L. established complex preparation, performed cryo-EM data analysis, model building, and refinement. C.C. assisted with cryo-EM analysis. C.P., P.-C.L. and K.N. analysed the structure and wrote the manuscript. C.P. and P.-C.L. designed the project and K.N. supervised the

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-

Supplementary information is available for this paper at https://doi.org/ 10.1038/s41586-018-0323-8

Reprints and permissions information is available at http://www.nature.com/ reprints.

Correspondence and requests for materials should be addressed to C.P., P.-C.L. or K.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Prespliceosome preparation and purification. To obtain the prespliceosome A-complex for structural studies, we prepared yeast S. cerevisiae containing a genomic TAPS affinity tag on the U2 snRNP subunit Hsh155, essentially as described³¹. Yeast cells were grown in a 120-l fermenter, and splicing extract was prepared using the liquid-nitrogen method, essentially as described³². Capped UBC4 pre-mRNA containing a point mutation (U > A) two nucleotides upstream of the branch point adenosine and three MS2 stem loops at the 3' end was produced by in vitro transcription^{9,33}. The RNA product was labelled with Cy5 at its 3' end to monitor complex purification³⁴. The pre-mRNA substrate was bound to the MS2-MBP fusion protein and added to an in vitro splicing reaction carried out for 90 min at 23 °C, essentially as described³³. The reaction mixture was then centrifuged through a 40% glycerol cushion in buffer A (20 mM HEPES (pH 7.9), 50 mM KCl, 0.2 mM EDTA, 1 mM dithiothreitol (DTT), 0.04% NP-40). The cushion was diluted with buffer A containing 1% glycerol, and applied to amylose resin (NEB) pre-washed with buffer B (20 mM HEPES (pH 7.9), 75 mM KCl, 5% glycerol, 0.2 mM EDTA, 1 mM DTT, 0.03% NP-40). After 12 h incubation at 4°C, the resin was washed with buffer B and eluted in buffer B containing 50 mM KCl and 12 mM maltose. Fractions containing A-complex were pooled and applied to Strep-Tactin resin (GE Healthcare), pre-washed with buffer B, and incubated for 4h at 4°C. The resin was washed with buffer B containing 2 mM MgCl₂, and eluted with buffer B containing 50 mM KCl, 2.5 mM desthiobiotin, and 2 mM MgCl₂. The A-complex fractions were pooled and crosslinked using 1.1 mM BS3 (Sigma) on ice for 1 h, and subsequently quenched with 50 mM ammonium bicarbonate. The sample was concentrated to \sim 0.4 mg ml⁻¹ and immediately used for EM sample preparation. Mass spectrometry (data not shown), indicated that homogenous A-complex was purified, containing sub-stoichiometric amounts of Prp5 (Extended Data Fig. 1b). The splicing assay in Extended Data Fig. 1a was carried out as for A-complex purification, but in a volume of 25 µl and in the absence of MS2-MBP fusion protein, and was visualized after 30 min of splicing at 23 °C on a denaturing 14% polyacrylamide TBE gel with a Typhoon scanner (GE Healthcare). Electron microscopy. For cryo-EM analysis the A-complex sample was applied to R2/2 holey carbon grids (Quantifoil), precoated with a 5–7-nm homemade carbon film. Grids were glow-discharged for 20 s before deposition of $2.5\,\mu l$ sample $(\sim 0.4 \,\mathrm{mg \, ml^{-1}})$, and subsequently blotted for 2–3.5 s and vitrified by plunging into liquid ethane with a Vitrobot Mark III (FEI) operated at 4°C and 100% humidity. Cryo-EM data were acquired on three separate FEI Titan Krios microscopes (datasets one to three) operated in EFTEM mode at 300 keV, each equipped with a K2 Summit direct detector (Gatan) and a GIF Quantum energy filter (slit width of 20 eV, Gatan). Datasets one and three were recorded using 'Krios 1' and 'Krios 2' at the MRC-LMB, respectively, and dataset three using 'Krios 2' at the Astbury Biostructure Laboratory (University of Leeds). For dataset one, 5,935 movies were acquired using EPU (FEI) with a defocus range of $-0.4 \mu m$ to $-4.4 \mu m$ at a nominal magnification of $105,000 \times (1.13 \text{ Å pixel}^{-1})$. The camera was operated in 'counting' mode with a total exposure time of 13 s fractionated into 20 frames, a dose rate of $4.25\,\mathrm{e^-}$ pixel⁻¹ s⁻¹, and a total dose of $43\,\mathrm{e^-}$ Å⁻² per movie. Dataset two was collected in the same manner, except that 727 movies were recorded using SerialEM³⁵, at a nominal magnification of $105,000 \times (1.14 \,\text{Å pixel}^{-1})$, a total exposure time of 8 s fractionated into 20 frames, a dose rate of $4.33 \,\mathrm{e^-}$ pixel⁻¹ s⁻¹ and a total dose of $27 e^- \text{ Å}^{-2}$ per movie. Dataset three was collected with EPU (FEI) similar to dataset one, except that 2,745 movies were collected at a nominal magnification of $130,000 \times (1.07 \,\text{Å pixel}^{-1})$, a total exposure time of 8 s fractionated into 20 frames, a dose rate of $7.94e^-$ pixel⁻¹ s⁻¹ and a total dose of $56e^-$ Å⁻² per movie.

Image processing. Movies were aligned using MOTIONCOR2 $^{\bar{3}6}$ with 5×5 patches and applying a theoretical dose-weighting model to individual frames. Contrast transfer function (CTF) parameters were estimated using Gctf^{S7}. Resolution is reported on the basis of the gold-standard Fourier shell correlation (FSC) (0.143 criterion) as described 38 and B-factors were determined and applied automatically in RELION 2.139,40. Particles from dataset one were automatically picked using Gautomatch (K. Zhang) and screened manually, and were then extracted in RELION with a 5,602 pixel box size and pre-processed. Particles from datasets two and three were picked and pre-processed in the same way, and were then rescaled to the pixel size of dataset one (1.13 Å pixel⁻¹) in RELION 2.1 by Fourier cropping during particle extraction with a 5,602 pixel box. For rescaling, we first calculated 3D refinements in RELION 2.1 for each dataset (one to three) and performed real space correlation fits in UCSF Chimera to identify scaling factors for datasets two and three relative to dataset one. Because the absolute magnification values differed slightly for the different microscopes, we re-determined the CTF values for datasets two and three using the new pixel sizes with Gctf³⁷, and then re-extracted and rescaled the particles to the 5,602 pixel box. Combining datasets one to three yielded a total dataset of 406,272 particles that were used for subsequent processing.

The first 22,319 particles from dataset one were used to generate an ab initio 3D reference for the A-complex using default parameters and three classes in cryoSPARC⁴¹ (Extended Data Fig. 2a). The complete dataset (one to three) was

subjected to a 'heterogeneous' (multi-reference) refinement in cryoSPARC using default parameters and four classes: the ab initio A-complex reference and three 'junk' references (Extended Data Fig. 2a; round 1). Class one contained 153,570 particles (37.8%, percentage of particles form the full dataset) and was used for a 3D refinement in RELION 2.1 with a soft mask in the shape of the A-complex. This yielded a density (map A1) with an overall resolution of 4.9 Å and a B-factor of $-188 \, \text{Å}^2$, comprising U1 snRNP and the U2 snRNP 3' region (Extended Data Figs. 1e, d, 2, 9). To improve the U1 snRNP density, we prepared a soft mask enveloping the U1 snRNP with the volume eraser in UCSF Chimera 42 and RELION 2.1^{39,40}. This enabled the focused refinement of the U1 snRNP (map A2) from the same 153,570 particles to an overall resolution of 4.0 Å and a B-factor of $-146 \,\text{Å}^2$ (Extended Data Figs. 1e, d, 2, 9). In the A-complex the U2 snRNP 5' region is flexible relative to the U1 and the U2 3' region (Extended Data Fig. 2). To position the U2 snRNP 5' region in the A-complex, we used a soft mask surrounding the U2 5' region and carried out 3D classification without image alignment with six classes (Extended Data Fig. 2a; round 2). This revealed a class with defined U2 5' region from 19,937 particles (4.9%) that could be refined to an overall resolution of 10.4 Å (Extended Data Figs. 2, 9). Local resolution was estimated using ResMap⁴³ (Extended Data Fig. 2d, e).

Structural modelling. We prepared a composite model of the A-complex by combining the A1-3 densities (Extended Data Fig. 1e, f). Model building was carried out in COOT⁴⁴. The U1 snRNP coordinates were refined into the sharpened A2 density in PHENIX $^{\!45}$ using the phenix.real_space_refine routine, and applying secondary structure, rotamer, nucleic acid and metal ion restraints. Homology models for yeast Yhc1, Snp1, and Mud1 were generated using MODELLER⁴⁶ from the human U1 snRNP crystal structures¹⁰ (PDB ID 4PJO, 4PKD) and were fitted and manually adjusted in the A2 map. The yeast B-complex U5 Sm ring model was used as the initial model for the U1 Sm ring, and was manually adjusted in the A2 density. Initial models for Prp39 and Prp42 were generated by I-TASSER⁴⁷ and were subsequently adjusted and extended manually. The Prp39 N-terminal residues 47-339 were modelled as poly-alanine owing to a lower local resolution of \sim 5–6 Å (Extended Data Figs. 2d, e, 3c). Snu56, the Yhc1 C terminus, the Snu71 N terminus were modelled de novo; Yhc1 residues 48-82 and 135-142 were modelled as poly-alanine. To build the Luc7 model a C₃H-type ZnF (from PDB ID 1RGO) for ZnF1 and a C₂H₂-type ZnF (from Yhc1) for ZnF2 were used to guide modelling in the A2 density, with a local resolution of 4-5 Å (Extended Data Fig. 3c). The helices connecting Luc7 ZnF1 and ZnF2 (α 5–7) were modelled as polyalanine, and were assigned on the basis of density connectivity. The U1 snRNP protein model is in excellent agreement with biochemical and protein crosslinking results¹². The U1 snRNA model was generated on the basis of similarity to the U1 snRNA in the human U1 snRNP crystal structures (PDB ID 3CW1, 4PJO, 4PKD) and according to the yeast U1 snRNA secondary structure prediction⁴⁸. All basepairing U1 snRNA regions (helix H; SL1; SL2-1 and -2; SL3-1, -2, -3, -4, -5 and -6), except for the SL3-7 and the tip of SL3-3, were modelled (Extended Data Fig. 3f, g). The human SL1 loop (PDB ID 4PKD) was rigid-body-fitted together with the homology model of the yeast Snp1 (described above), and the human U1 snRNA sequence was replaced with that of yeast. The loops connecting SL2-1 to SL2-2 as well as SL3-3 to SL3-4 and SL3-4 to SL3-5 and the tips of SL2-2, SL3-3, -4 and -5 $\,$ were not built, owing to a lower local resolution (\sim 4.5 Å). The location of a region of U1 snRNA SL3-7 was modelled as a phosphate backbone only and may correspond to the sequence surrounding residues 378-391 and 428-440. The U1-5'SS was modelled de novo, and the UBC4 pre-mRNA contained 12 nucleotides, ten from the intron (+1 to +10) and two from the 5' exon (-1 to -2).

The U2 snRNP 3′ region (U2 3′ domain and SF3a subcomplexes) from the yeast B-complex structure (PDB ID 5NRL) were fitted into the A1 density using UCSF Chimera⁴², and the positions of Lea1, Msl1 and U2 snRNA residues 139–1169 were adjusted as a rigid body in COOT⁴⁴. The U2 snRNP 5′ region from the yeast B-complex structure (PDB ID 5NRL) was fitted into the A3 density in UCSF Chimera. This provided an excellent fit, suggesting that the U2 5′ region structure is not changed substantially from that observed in the yeast B-complex⁸. To generate the complete A-complex model, the refined U1 snRNP model and the U2 snRNP 3′ region were fitted into the A3 density in UCSF Chimera, together with the fitted U2 snRNP 5′ region. The final model comprises 34 proteins, U1 and U2 snRNAs, and the pre-mRNA substrate.

To generate the alternative pre-B-complex model shown in Fig. 3, we modified and combined structural models using COOT⁴⁴, on the basis of structural and biochemical data from yeast and human systems^{8,25,28}. We first superimposed our A-complex structure on the yeast B-complex structure⁸ using the U2 SF3b-containing domain. The free human tri-snRNP structure (PDB ID 3JCR), which may resemble the pre-B conformation^{7,25}, was used to model the yeast tri-snRNP in the pre-B-complex conformation. We first removed the B-complex proteins from the yeast B-complex structure, because these are absent in the purified human pre-B-complex²⁸. Human pre-B instead contained the PRP28 helicase and SAD1, and we therefore placed crystal structures of the yeast Prp28 helicase⁴⁹ (PDB ID

4W7S) and yeast Sad1⁵⁰ (PDB ID 4MSX) in their human tri-snRNP locations²⁵. We then positioned the U4 Sm ring and Brr2 as in the human tri-snRNP structure, in which the Brr2 PWI domain makes a conserved contact with Sad1⁵¹. We removed a Snu66 peptide bound to Brr2 from the model, because its binding at this site is uncertain in the pre-B-complex conformation. Several minor differences remain between the free human tri-snRNP structure²⁵ and the pre-B-complex model, and these were not modelled. The final pre-B model contained only minor clashes, and one observed clash between the highly flexible Prp28 RecA-2 lobe²⁵ and the flexible U6 snRNA 5' stem loop^{8,26} could be resolved by a minor repositioning of either domain. The final pre-B model comprises 66 proteins, five snRNAs, the pre-mRNA substrate, and has a combined molecular mass of $\sim 3.1\,\mathrm{MDa}$.

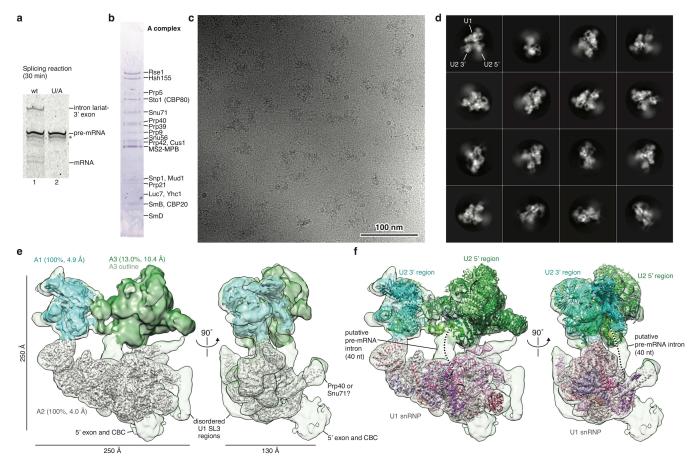
Figures were generated with PyMol (https://www.pymol.org) and UCSF Chimera.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Three-dimensional cryo-EM density maps A1, A2 and A3 have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-4363, EMD-4364 and EMD-4365, respectively. The coordinate file of the A-complex has been deposited in the Protein Data Bank under the accession number 6G90.

- 31. Nguyen, T. H. D. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature **523**, 47–52 (2015).
- Umen, J. G. & Guthrie, C. A novel role for a U5 snRNP protein in 3' splice site selection. Genes Dev. 9, 855–868 (1995).
- Galej, W. P. et al. Cryo-EM structure of the spliceosome immediately after branching. Nature 537, 197–201 (2016).
- Wang, H., Ach, R. A. & Curry, B. Direct and sensitive miRNA profiling from low-input total RNA. RNA 13, 151–159 (2007).
- Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. 152, 36–51 (2005).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017).
- Zhang, K. Gctf: Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12 (2016).
- Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854 (2012).
- Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. eLife 5, e18722 (2016).

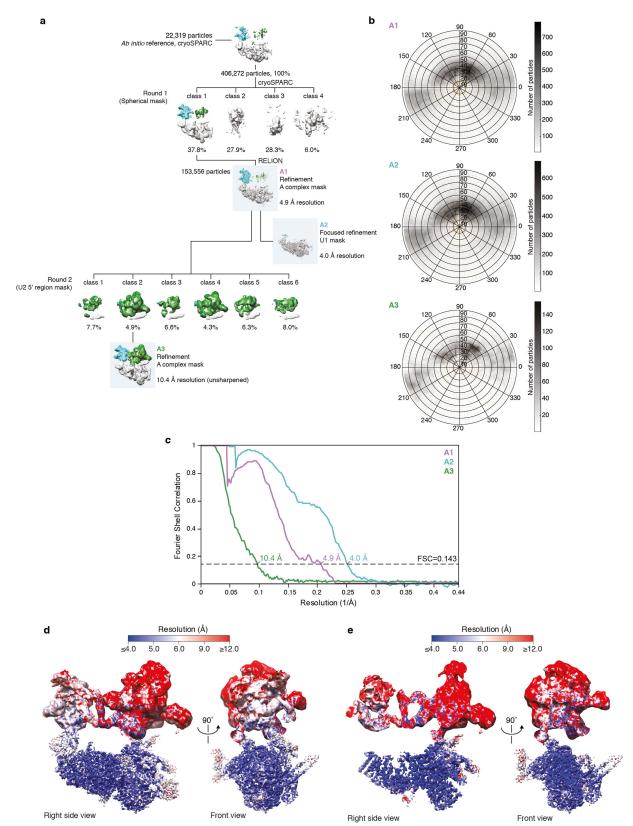
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017).
- 42. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65 (2014).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66, 213–221 (2010).
- 46. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 5.6.1–5.6.32 (2014).
- Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8 (2015).
- Kretzner, L., Krol, A. & Rosbash, M. Saccharomyces cerevisiae U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. Proc. Natl Acad. Sci. USA 87, 851–855 (1990).
- Jacewicz, A., Schwer, B., Smith, P. & Shuman, S. Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. *Nucleic Acids Res.* 42, 12885–12898 (2014).
- Hadjivassiliou, H., Rosenberg, O. S. & Guthrie, C. The crystal structure of S. cerevisiae Sad1, a catalytically inactive deubiquitinase that is broadly required for pre-mRNA splicing. RNA 20, 656–669 (2014).
- Absmeier, E., Santos, K. F. & Wahl, M. C. Functions and regulation of the Brr2 RNA helicase during splicing. Cell Cycle 15, 3362–3377 (2016).
- Ester, C. & Uetz, P. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. BMC Biochem. 9, 29 (2008).
- 53. Kimura, E. et al. Serine-arginine-rich nuclear protein Luc7l regulates myogenesis in mice. *Gene* **341**, 41–47 (2004).
- 54. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911 (2016).
- Wilkinson, M. E. et al. Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. Science 358, 1283–1288 (2017).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, W320–W324 (2014).
- Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).
- Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* 41, W349–W357 (2013).



Extended Data Fig. 1 | Biochemical characterization and cryo-EM of the prespliceosome A-complex. a, Mutation of the *UBC4* pre-mRNA branch point sequence (UACUAAC to UACAAC, in which A is the branch-point adenosine and A is the mutated nucleotide) stalls splicing before the first step, as described⁹. Splicing reactions were carried out for 30 min at 23 °C in yeast extract using wild-type (lane one) or mutant (U > A, lane two) pre-mRNA. This experiment was performed three times. The asterisk indicates a degradation product. For gel source data see Supplementary Fig. 1a. b, Protein analysis of purified A-complex (SDS-PAGE stained with Coomassie blue). The U2-associated Prp5 protein is sub-stoichiometric and not observed in the A-complex structure. The purification and analysis of protein compositions were performed at least five times with similar results. For gel source data see Supplementary Fig. 1b. c, Cryo-EM micrograph of the A-complex. Scale bar, 100 nm. d, 2D class

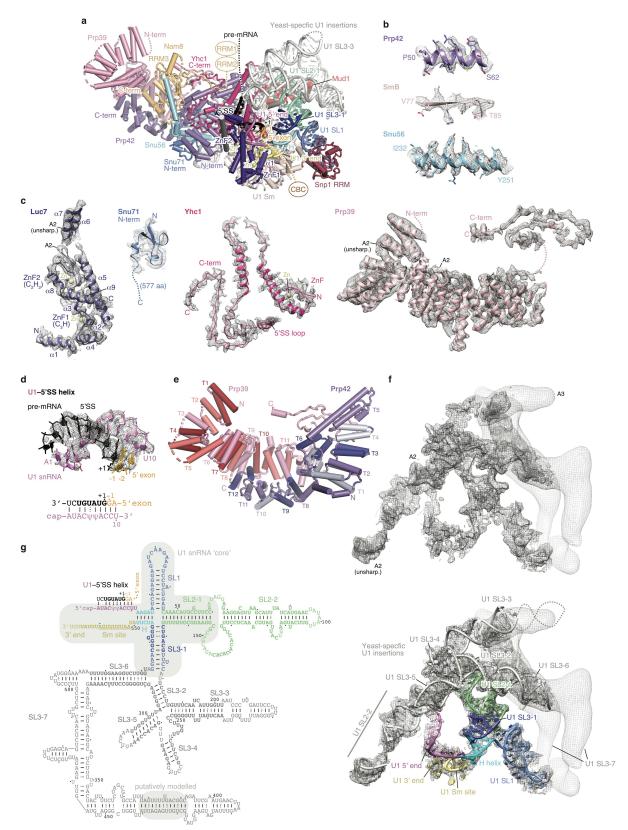
averages of the A-complex were determined in RELION 2.1^{39,40}, and reveal a bipartite architecture, comprising the U1 snRNP and the U2 snRNP 3' and 5' regions, respectively. **e**, Composite cryo-EM density of the A-complex shown in two orthogonal views (compare to Fig. 1). The respective densities used for modelling the U1 snRNP (A2, grey), the U2 3' region (A1, cyan), and the U2 5' region (A3, green) are coloured and superimposed on a transparent outline of the full A3 map (Methods). The overall resolution of each map as well as the percentage from the cleaned dataset of 153,556 particles are shown in parentheses. Non-modelled regions are indicated and putatively assigned. **f**. Composite cryo-EM density with the final A-complex model superimposed in a cartoon representation. The path of 40 nucleotides of the disordered *UBC4* pre-mRNA intron are indicated. A-complex components are coloured as in Fig. 1. Views as in **e**.





Extended Data Fig. 2 | **Cryo-EM image classification and refinement. a**, Image processing workflow for analysis of the A-complex cryo-EM dataset (see 'Image processing' in Methods). To visualize differences between the reconstructions, the U1 snRNP (grey), U2 3' (cyan) and U2 5' regions (green) are coloured. For each round of three-dimensional classification, the percentage of the data and the type of soft-edged mask are indicated. The type of mask and overall resolution are indicated for

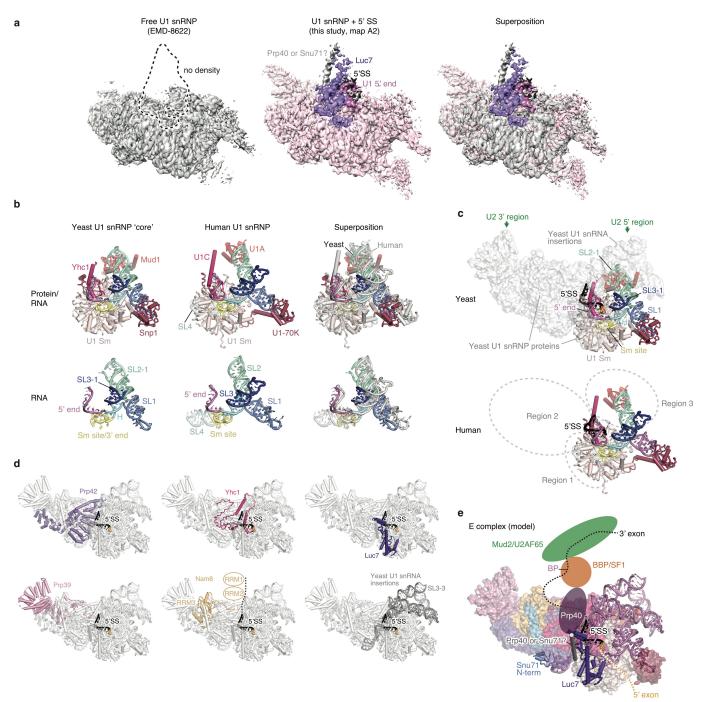
each 3D refinement (blue box). **b**, Orientation distribution plots for all particles that contribute to the respective A1, A2, and A3 cryo-EM reconstructions. **c**, Gold-standard Fourier shell correlation (FSC = 0.143) of the respective A1, A2 and A3 cryo-EM reconstructions. **d**, Two views of the composite A-complex cryo-EM density (maps A1, A2 and A3) coloured by local resolution as determined by ResMap 43 . **e**, As panel **d**, but for a central slice through the composite A-complex cryo-EM map.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Details of the U1 snRNP. a, U1 snRNP structure with subunits coloured as in Fig. 1, except for Nam8 (orange), Snu56 (light blue), Snu71 (blue), Luc7 (dark purple), Mud1 (red) and the U1 snRNA (various). The pre-mRNA nucleotides are labelled relative to the first nucleotide (+1) of the intron. The Nam8 RRM1 and RRM2 domains are flexible and project downstream of the 5'SS. The protein attributed to Luc7 in the free U1 snRNP structure¹² was re-assigned to Snu71. C-term, C terminus; N-term, N terminus; SL, stem loop. In the structure we do not observe any evidence that the C-terminal tails of SmB, SmD1, and SmD3 interact with the 5'SS, consistent with their absence in the human 5'SSminimal U1 snRNP crystal structure¹⁰. **b**, Representative regions of the sharpened U1 snRNP density determined at 4 Å resolution (map A2) are superimposed on the refined coordinate model. The density reveals sidechain details, and here segments from the Prp42 N terminus (TPR repeat 1), the Sm ring subunit SmB, and the Snu56 α -helical domain are shown. c, The A2 cryo-EM density is shown superimposed on the coordinate models of a selection of U1 snRNP proteins: Luc7, Snu71, Yhc1 and Prp39. In the structure most of Snu71 is disordered, except for a small N-terminal domain (residues 2-43) that binds between the Prp42 N terminus and the Snu56 KH-like fold, consistent with protein crosslinking¹². Functional

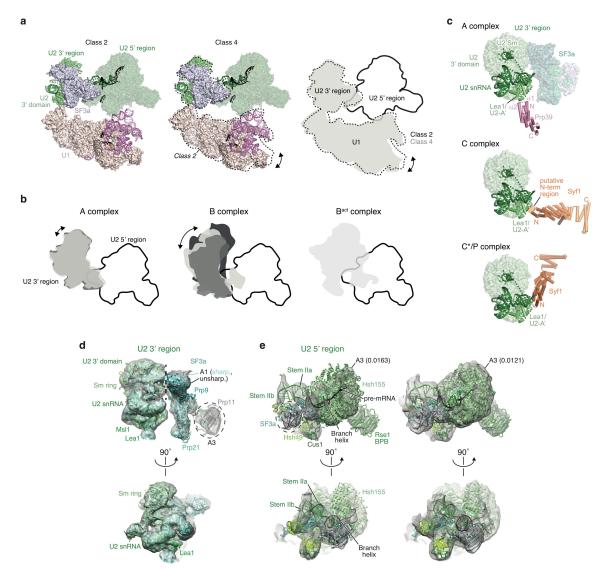
regions and disordered domains are indicated. d, The U1 snRNA-pre-mRNA 5' splice site (U1-5'SS) model is superimposed on its cryo-EM density (map A2). A secondary structure diagram of the U1-5'SS interaction is shown underneath the model. The register of the U1-5'SS is shifted by one nucleotide with respect to U1C (Yhc1) compared to the minimal human 5'SS-U1 snRNP crystal structure, owing to an additional nucleotide in the yeast U1 snRNA¹⁰ (U11). Lines indicate Watson-Crick base pairs and dots indicate pseudouridine (ψ)-containing base pairs. **e**, The Prp39–Prp42 heterodimer is coloured to indicate each of their respective TPR repeats. f, Cryo-EM density of U1 snRNA from maps A2 (dark grey) and A3 (light grey) without (top) and with the superimposed coordinate model of yeast U1 snRNA (bottom). The model is labelled and coloured according to functional regions of U1 snRNA (5' end, pink; H helix, cyan; SL1, dark blue; SL2-1, green; SL3-1, light blue; SL2-2 and SL3-2 to -7, grey; 3'end and Sm site, yellow). g, Secondary-structure diagram of U1 snRNA. Bold letters indicate residues included in the model, lines indicate Watson-Crick base pairs, and dots G-U wobble and pseudouridinecontaining base pairs. Compare to e. The conserved U1 snRNA 'core' is outlined with a grey box. The region of the putative phosphate backbone model of part of the U1 SL3-7 region is indicated with a grey box.



Extended Data Fig. 4 | See next page for caption.

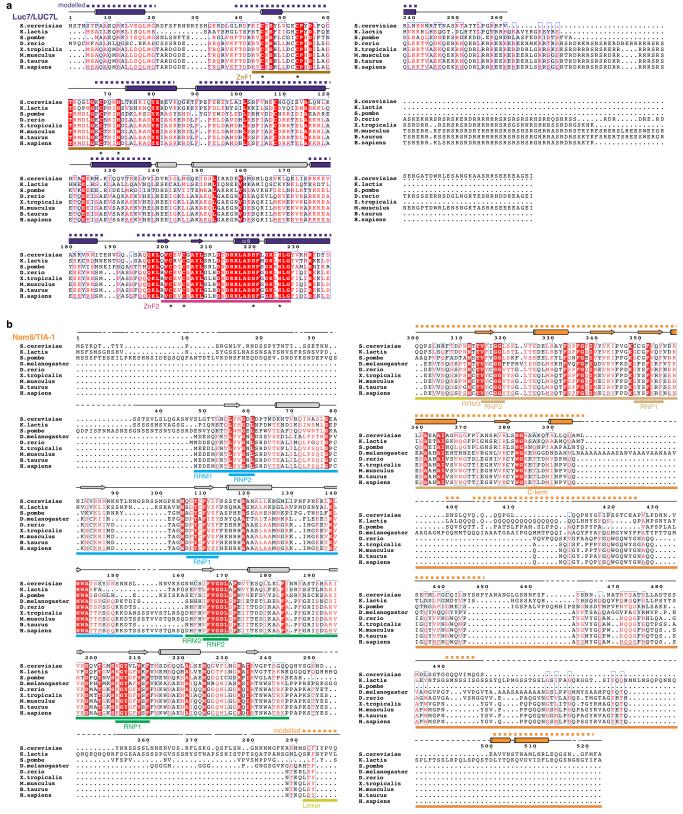
Extended Data Fig. 4 | Comparisons of yeast and human U1 snRNPs and implications for alternative splicing. a, Formation of the U1-5'SS helix induces stable binding of Luc7. In the absence of a pre-mRNA 5'SS in the free U1 snRNP density (left, EMD-8622), Luc7 and the U1 5' end are disordered. Upon 5'SS recognition at the U1 5' end (centre, map A2), Luc7 becomes ordered and stabilizes the U1-5'SS interaction, suggesting a mechanism for the selection of weak 5'SS sequences. The free U1 snRNP and the 5'SS-bound (map A2) cryo-EM densities are superimposed on the right. Although the long α -helical density next to Luc7 cannot be assigned with confidence, protein-protein crosslinking data¹² and protein secondary structure prediction are consistent with the presence of either Prp40 or Snu71. On the basis of additional biochemical data on the interaction between the α -helical Prp40 FF1 domain and Luc7 ZnF2⁵², we would speculate that the Prp40 FF1 domain is the most likely candidate for this density. b, Comparison of the yeast U1 snRNP 'core' with the human U1 snRNP crystal structure (PDB ID 3CW1). Protein and RNA (top) and RNA only (bottom) are shown side by side (left and centre) and superimposed by a global alignment in PyMOL (right). Coloured as in Extended Data Fig. 3a. c, The yeast U1 snRNP model suggests regulatory mechanisms for human alternative splicing factors. The human homologues of the peripheral yeast U1 proteins may function through stabilization of the U1-5'SS interaction (region 1), of the U1-U2 3' region

interface (region 2), or the U1-U2 5' interface (region 3). The yeast U1 snRNP 'core' is shown superimposed on a surface representation of the U1 snRNP model (top), compared with the similarly coloured human U1 snRNP (below). Interaction sites with the U2 snRNP are labelled (top). d, The location of yeast U1 snRNP components with homology to human splicing factors are indicated in the U1 snRNP structure. The Prp39-Prp42 heterodimer (human PRPF39 homodimer), Nam8¹⁸ (human TIA-1 and TIA-R), Luc7⁵³ (human LUC7L1-3), and the Yhc1 C terminus (human U1C) have clear counterparts in the human system. The yeast-specific U1 snRNA insertions may be replaced in the human system by alternative splicing factors that modulate interactions with the U2 5' region. e, Model of the yeast E complex on the basis of the U1 snRNP structure and biochemical data²². Luc7, Snu71 and Prp40 form a heterotrimer in vitro⁵², and their interacting regions may be located near unassigned density (compare to Extended Data Fig. 1e) at the tip of an unassigned 40-residue α -helix next to Luc7 ZnF2. This helix is likely to belong to the U1 subunit Snu71 or Prp40, consistent with protein crosslinking¹² and protein secondary-structure prediction. Prp40 could then bind the yeast branch point-binding protein (BBP, human SF1), which in turn interacts with Mud2 (human U2AF65) to tether the pre-mRNA branch-point sequence in the E complex²².



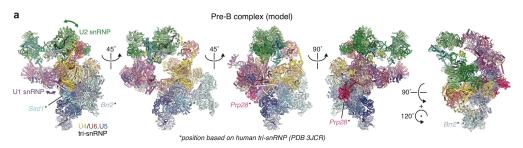
Extended Data Fig. 5 | **Conformational flexibility of the U2 snRNP. a**, Two defined positions of the U1 snRNP-U2 3' region could be identified relative to the U2 5' region. A-complex models were fitted into class two and four from round two of the 3D image classification (compare Extended Data Fig. 2a). The classes are aligned via their U2 5' region, illustrating their relative flexibility. **b**, Cartoon schematic of observed positions of the U2 3' region relative to the U2 5' region in the A-complex (left), B-complex⁸ (centre), and activated B-complex (B^{act}) (right, modelled from previously published work⁵⁴). Although in the B-complex the U2 3' region is free, in the A- and B^{act}-complexes the position of the U2 3' region is influenced by interactions with Prp39 as well as Syf1 and Clf1, respectively. **c**, The U2 snRNP subunit Lea1 (human U2A') aids to position the U2 snRNP 3' domain in different spliceosome states. In our A-complex structure, the Prp39 TPR repeat T1 contacts the helical C terminus of Lea1. In the yeast C-complex structure, the non-modelled

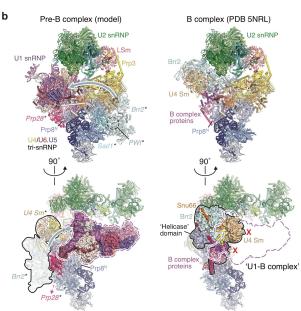
density for the Syf1 N terminus binds a neighbouring but non-overlapping surface of Lea1 (PDB ID 5LJ5). In the C*/P-complex⁵⁵ (PDB ID 6EXN), the Syf1 N terminus binds yet another Lea1 surface and the U2 3′ domain is repositioned relative to its C-complex location. Together, this suggests that the Lea1 provides multiple interfaces that can be used to position the U2 3′ domain in different spliceosomal complexes. d, Fit of the U2 3′ region coordinate model to the A1 cryo-EM density. The dashed black lineseparates the U2 3′ domain (Sm ring, Msl1 and Lea1 subunits and U2 snRNA, left) and the SF3a subcomplex (Prp9, Prp11 and Prp21, right). Two orthogonal views are shown (Supplementary Video 2). e, Fit of the U2 5′ region coordinate model to the A3 cryo-EM density. A density consistent with the U2 snRNA stem IIa/b and the branch helix is observed. Two density thresholds are shown side by side (left, 0.0163; right, 0.0121), and orthogonal views are shown underneath (Supplementary Video 2).



Extended Data Fig. 6 | Luc7 and Nam8 sequence alignments. a, The Luc7 (human LUC7-like) amino-acid sequence alignment comparing *S. cerevisiae, Kluyveromyces lactis, Schizosaccharomyces pombe, Danio rerio, Xenopus tropicalis, Mus musculus, Bos taurus* and *Homo sapiens* was generated with Clustal Omega and visualized with ESPript 3^{56,57}. For the human sequence, LUC7L1 was used. Secondary structure elements are indicated above the sequence and derive from the A-complex structure

(purple) or PSIPRED⁵⁸ secondary structure prediction (grey). Modelled regions (dashed line) and the Zn-coordinating residues of ZnF1 and ZnF2 (asterisks) are indicated. Invariant or conserved residues are highlighted with a red box or red letter font, respectively. **b**. As in panel **a** but for Nam8 (human TIA-1) comparing *S. cerevisiae*, *K. lactis*, *S. pombe*, *Drosophila melanogaster*, *D. rerio*, *X. tropicalis*, *M. musculus*, *B. taurus*, and *H. sapiens* amino acid sequences.

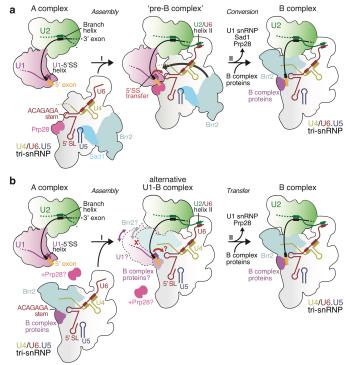




Extended Data Fig. 7 | Details of the pre-B-complex model.

a, Multiple views of the pre-B-complex model, generated by combining functional and structural data from yeast and human systems^{8,25}. The mobility of the U1 snRNP relative to the U2 snRNP in the A-complex (this study) as well as of the U2 snRNP relative to tri-snRNP in the B-complex structure⁸ are indicated (left). The pre-B model contained only minor clashes, and a clash between the highly flexible Prp28 C-terminal RecA-2 lobe (from the human tri-snRNP²⁵) and the highly flexible U6 snRNA 5' stem loop (from the yeast B-complex⁸) may be resolved by small movements of either domain. b, Structural comparisons of the yeast pre-B model (from this study) and the yeast B-complex structure (PDB ID 5NRL⁸) suggest the existence of a molecular checkpoint to couple 5'SS transfer to U1 snRNP release and formation of the activation-competent B-complex. In the pre-B model (left) Sad1 tethers Brr2 through its

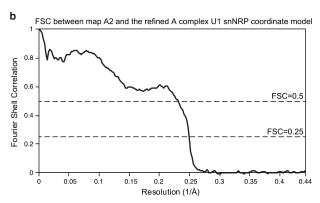
interaction with the conserved Brr2 PWI domain⁵¹, and the U1 snRNP and its U1–5'SS helix are positioned near the U6 ACAGAGA region and the helicase Prp28. Subsequent to Prp28-mediated 5'SS transfer, Brr2 is repositioned onto its U4 snRNA substrate, guided by the B-complex-specific proteins (right). In this conformation the Brr2 helicase and its associated factors would clash with the U1 snRNP, consistent with U1 snRNP destabilization and release yeast and human B-complexes^{7,8}. Brr2 is now ready to initiate spliceosome activation and formation of the active site in the B^{act}-complex. Regions that are changed between pre-B- and B-complex models (black outline) and the clash between the Brr2-containing 'helicase' domain and the U1 snRNP in B-complex (red X) are indicated. The lower right panel would conform to the alternative 'U1-B-complex' model.



Extended Data Fig. 8 | Model for early splicing events. a, Cartoon schematic of proposed early splicing events, detailing (i) assembly of the pre-B-complex spliceosome from the A-complex and the U4/U6.U5 trisnRNP and (ii) the subsequent conversion to the pre-catalytic B-complex spliceosome. In the pre-B model the mobile U1 snRNP is next to Prp28, which is bound at the Prp8^N domain. To initiate 5'SS transfer, Prp28 could clamp the pre-mRNA at, or next to, the U1-5'SS helix to destabilize it and to hand over the 5'SS to the U6 ACAGAGA region of tri-snRNP, consistent with protein-RNA crosslinks³⁰. Transfer of the 5'SS may induce the binding of the B-complex proteins to replace Prp28 at the Prp8^N domain and induce the large movement of Brr2 to its B-complex location on U4 snRNA. The U1 snRNP, now loosely tethered to U2, may dissociate from the B-complex owing to the steric clash with the Brr2containing 'helicase' domain⁸ (Extended Data Fig. 7b). Consistent with this, the human pre-B-complex converts to a B-complex-like state in the presence of a 5'SS oligonucleotide, which coincides with U1 snRNP release²⁸. This model can explain how Brr2 is kept inactive to prevent premature U4/U6 duplex unwinding²⁶. The model thereby implies the existence of a molecular checkpoint, coupling 5'SS transfer from U1 to U6 snRNA with Brr2 helicase repositioning and U1 snRNP release to generate the activation-competent B-complex spliceosome. b, Cartoon schematic of an alternative model for spliceosome assembly and 5'SS transfer that relies only on the yeast A-complex (from this work), tri-snRNP^{26,29} and B-complex structures⁸. In this model the tri-snRNP that associates with the A-complex already contains the Brr2 helicase bound to the U4 snRNA substrate and the yeast B-complex proteins at the Prp8 N-terminal domain. The tri-snRNP then binds the A-complex (transition I, 'Assembly'), requiring a substantial readjustment to avoid a steric clash of the Brr2-containing 'helicase' domain and the U1 snRNP ('U1-Bcomplex'). The Prp28 helicase is then recruited to the U1 snRNP directly as the Prp28-binding site on the Prp8 N-terminal domain in human trisnRNP is occupied by B-complex proteins²⁵. Prp28 then disrupts the U1-5'SS helix, leading to 5'SS transfer (transition II, 'Transfer'). Similar to the 'pre-B-complex' assembly model in a, the U1 snRNP, now freed from the 5'SS, may then be released owing to a steric clash with the Brr2-containing 'helicase' domain. This model does not require Sad1. Compare to a.

a Cryo-EM data collection and refinement statistics of the A complex structure

	A1 (U2 3' region)	A2 (U1 snRNP)	A3 (U2 5' region)	
Data collection				
Particles	153,556	153,556	19,937	
Pixel Size (Å)	1.13	1.13	1.13	
Defocus range (µm)	−0.4 to −4.4	−0.4 to −4.4	−0.4 to −4.4	
Voltage (kV)	300	300	300	
Electron dose (e ⁻ A ⁻²)	27-56	27-56	27-56	
Reconstruction (RELION)				
Accuracy of rotations (°)	1.21	0.95	1.70	
Accuracy of translations (pixel)	1.47	0.92	1.91	
Resolution (Å)	4.9	4.0	10.4	
Map sharpening B-factor (Å ²)	-188	-146	0	
Model composition				
Non-hydrogen atoms	13,333	28,244	23,788	
Protein residues	1,408	2,803	2,765	
RNA bases	91	338	78	
Refinement (PHENIX)				
Map CC (around atoms)		0.738		
Rms deviations				
Bond lengths (Å)		0.016		
Bond angles (°)		2.02		
Validation				
Molprobity score		1.98		
All-atom clashscore		8.06		
Rotamer outliers (%)		0.34		
C-beta deviations		2		
Ramachandran plot				
Outliers (%)		0.5		
Allowed (%)		9.21		
Favoured (%)		90.29		
RNA validation		50.25		
Correct sugar puckers (%)		97.6		
Good backbone conformations (%)				
. ,		77.2		
Data Deposition	EMD 42/2	EMD 4274	EMD 4265	
EMDB ID	EMD-4363	EMD-4364	EMD-4365	
PDB ID for the complete model	6G90			



Extended Data Fig. 9 | Data collection, refinement statistics and validation. a, Cryo-EM data collection and refinement statistics of the A-complex structure. Maps A1 and A3 were used to position the U2

snRNP 3' and 5' regions, respectively. $\bf b.$ FSC between the A2 cryo-EM density and the refined A-complex U1 snRNP coordinate model.



$\textbf{Extended Data Table 1} \mid \textbf{Summary of the components modelled into the A-complex cryo-EM densities}$

					Modelling			
Sub-complexes	Protein/RNA	Total residues	M.W. (kDa)	Modelled residues	template (PDB ID)	Modelling	Chain ID	Human name
	Mud1	298	34.4	17-42; 62-81; 84-94 97-123; 134-148	4PKD	Docked	Α	U1A
	Snp1	300	34.4	5-55; 58-88; 94-204	4PJO, 4PKD	Docked and rebuilt	В	U1-70K
	Yhc1	231	27.0	2-59; 67-142; 153-195	4PJO, 4PKD	Docked, rebuilt, de novo	С	U1C
	Prp39	629	74.8	47-63; 66-85; 88-102; 108-119 124-136; 139-154; 160-172; 177-190; 193-208; 217-236; 250-266; 271-275; 276-286; 289-304; 307-321; 325-382; 388-553; 561-626		de novo	D	PRPF39
	Prp42	544	65.1	2-542		de novo	E	PRPF39
U1 snRNP	Nam8	523	57.0	292-400; 404-425; 434-449 491-497; 501-521		de novo	F	TIA-1
	Snu56	492	56.5	45-104; 109-170; 185-294		de novo	G	
	Luc7	261	30.2	5-20; 39-59; 67-84; 91-120 126-138; 175-187; 195-241		de novo	Н	LUC7L
	Snu71	620	71.4	2-43		de novo	J	RBM25
	SmB	196	22.4	2-63; 73-131	5NRL	Docked and adjusted	b	SmB
	SmD3	101	11.2	3-95	5NRL	Docked and adjusted	d	SmD3
	SmD1	146	16.3	1-73; 78-119	5NRL	Docked and adjusted	h	SmD1
	SmD2	110	12.9	8-108	5NRL	Docked and adjusted	i	SmD2
	SmE	96	9.7	8-63; 73-93	5NRL	Docked and adjusted	е	SmE
	SmF	86	10.4	12-84	5NRL	Docked and adjusted	f	SmF
	SmG	77	8.5	2-77	5NRL	Docked and adjusted	g	SmG
	U1 snRNA	568		1-61; 67-95; 103-112; 115-144; 52-173; 181-202; 236-258; 260-264 ;70-275; 280-287; 295-325; 378-394 424-440; 516-532; 538-564		Docked and de novo	1	
	Unknown			1-56		de novo	Х	
	Msl1	111	12.8	28-111	5NRL	Docked	Y	U2-B"
	Lea1	238	27.2	1-170	5NRL	Docked	W	U2-A'
	SmB	196	22.4	12-54; 76-102	5NRL	Docked	S	SmB
	SmD3	101	11.2	4-85	5NRL	Docked	V	SmD3
	SmD1	146	16.3	1-48; 78-101	5NRL	Docked	t	SmD1 SmD2
	SmD2 SmE	110 96	12.9 9.7	17-108 10-63; 71-93	5NRL 5NRL	Docked Docked	u w	SmE
	SmF	86	10.4	12-84	5NRL	Docked	x	SmF
	SmG	77	8.5	2-76	5NRL	Docked	x V	SmG
	Hsh155	971	110.0	132-149; 157-971	5NRL	Docked	0	SF3B1
	Rse1	1361	153.8	53-305; 323-571; 581-784; 814-890; 918-1265; 1292-1361	5NRL	Docked	P	SF3B3
J2 snRNP	Cus1	436	50.3	125-213; 239-353; 361-376	5NRL	Docked	Q	SF3B2
-	Hsh49	213	24.5	9-86; 106-144; 147-185; 189-203	5NRL	Docked	R	SF3B4
	Rds3	107	12.3	2-104	5NRL	Docked	S	SF3B14b
	Ysf3	85	10.0	2-84	5NRL	Docked	Z	SF3B5
	Prp9	530	63.0	1-97; 112-378; 407-478; 503-528	5NRL	Docked	Т	SF3A3
	Prp11	266	29.9	34-47; 51-105; 115-136; 149-253	5NRL	Docked	U	SF3A2
	Prp21	280	33.1	89-206; 220-228	5NRL	Docked	V	SF3A1
	U2 snRNA	1175	363.8	3-13; 30-73; 79-86; 108-122; 139-150; 1089-1109; 1115-1130;	5NRL	Docked	2	
				1138-1154; 1159-1169				

LETTER

PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling

Hsin-Yung Yen^{1,2}, Kin Kuan Hoi¹, Idlir Liko^{1,2}, George Hedger³, Michael R. Horrell³, Wanling Song³, Di Wu¹, Philipp Heine⁴, Tony Warne⁵, Yang Lee⁵, Byron Carpenter^{5,6}, Andreas Plückthun⁴, Christopher G. Tate⁵, Mark S. P. Sansom³* & Carol V. Robinson¹*

G-protein-coupled receptors (GPCRs) are involved in many physiological processes and are therefore key drug targets¹. Although detailed structural information is available for GPCRs, the effects of lipids on the receptors, and on downstream coupling of GPCRs to G proteins are largely unknown. Here we use native mass spectrometry to identify endogenous lipids bound to three class A GPCRs. We observed preferential binding of phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) over related lipids and confirm that the intracellular surface of the receptors contain hotspots for PtdIns(4,5)P₂ binding. Endogenous lipids were also observed bound directly to the trimeric $G\alpha_s\beta\gamma$ protein complex of the adenosine A_{2A} receptor (A_{2A}R) in the gas phase. Using engineered G α subunits (mini- $G\alpha_s$, mini- $G\alpha_i$ and mini- $G\alpha_{12}$)², we demonstrate that the complex of mini-G α_s with the β_1 adrenergic receptor (β_1AR) is stabilized by the binding of two PtdIns(4,5)P₂ molecules. By contrast, PtdIns(4,5)P₂ does not stabilize coupling between β₁AR and other $G\alpha$ subunits (mini- $G\alpha_i$ or mini- $G\alpha_{12}$) or a high-affinity nanobody. Other endogenous lipids that bind to these receptors have no effect on coupling, highlighting the specificity of PtdIns(4,5) P2. Calculations of potential of mean force and increased GTP turnover by the activated neurotensin receptor when coupled to trimeric $G\alpha_i\beta\gamma$ complex in the presence of PtdIns(4,5)P₂ provide further evidence for a specific effect of PtdIns(4,5)P₂ on coupling. We identify key residues on cognate $G\alpha$ subunits through which PtdIns(4,5)P₂ forms bridging interactions with basic residues on class A GPCRs. These modulating effects of lipids on receptors suggest consequences for understanding function, G-protein selectivity and drug targeting of class A GPCRs.

The emerging view from biophysical studies of GPCRs is that they exist as ensembles of discrete conformations that can be influenced by ligands, regulatory proteins, pH, ions and, potentially, lipid molecules³. The complex roles of these conformational ensembles in signalling pathways are further compounded by the combinatorial effects of the multiple distinct heterotrimeric complexes formed from 21 G α , 6 G β and 12 G γ subunits. Investigating the relationship between GPCRs, small molecule modulators and numerous binding partners is therefore challenging, owing to the difficulty of observing the complexity of these interactions directly. A previous study characterized interactions of lipids with the β_2 adrenergic receptor (β_2AR) in high-density lipoparticles⁴ to which phospholipids were added exogenously, but did not address the selectivity and effects of different phosphatidylinositol (PI) phosphate lipids on coupling with downstream effectors. In this study, we develop and apply high-resolution native mass spectrometry to interrogate endogenous lipid-receptor interactions^{5,6} of three class A GPCRs: the β_1 adrenergic receptor $(\beta_1 AR)$, the adenosine A_{2A} receptor $(A_{2A}R)$, and neurotensin receptor 1 (NTSR1). We reveal effects of PtdIns(4,5)P₂ that stabilize these receptors in active states, increase GTPase activity and enhance selectivity of coupling to G proteins.

First, we considered the endogenous lipids that bind directly to β_1AR and the stabilized NTSR1(HTGH4- Δ IC3B), which were expressed in and purified from insect cells and Escherichia coli, respectively. Peaks corresponding to lipid adducts were observed for β_1AR and for NTSR1 (Fig. 1a and Extended Data Fig. 1a). Collisional dissociation of protein-lipid complexes allowed us to identify two major classes of lipids bound to β_1AR , the phosphatidylserines (PS) (34:2 and 36:2) and PI phosphates (42:5), as well as phosphatidic acid (PA) (36:2), which bound to NTSR1 (Extended Data Fig. 1b, c and Extended Data Table 1). To investigate this selectivity, we incubated NTSR1 with PA and other anionic lipids (PS and PI), a zwitterionic lipid (phosphatidylcholine (PC)), and a neutral lipid (diacylglycerol (DAG)). Analysis of the resulting native mass spectra show that NTSR1 interacts preferentially with PA, PS and PI (Extended Data Fig. 2a-e). We did not observe apparent binding of phosphatidylglycerol (PG) to NTSR1, although PG has been reported to increase G-protein activation of NTSR1 in a nanodisc8. It is possible that PG affects the local net charge at the receptor-lipid interface. Similarly, β_1AR , when incubated with detergent-solubilised PS (16:0–18:1) or phosphatidylinositol-4-phosphate (PtdIns(4)P) (18:1-18:1), showed higher affinity towards PtdIns(4)P than to PS (Fig. 1a and Extended Data Fig. 2f, g).

To probe the selectivity of different PI derivatives we incubated β_1AR with equimolar ratios of PI, PtdIns(4)P, phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) and phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃), all containing the same acyl chains (18:1–18:1). Plotting intensity of peaks corresponding to lipid-bound states in the mass spectrum, relative to those of the apo protein, showed that PtdIns(4,5)P₂ had a higher affinity than PtdIns(4)P for β_1AR (Fig. 1b). In the case of PtdIns(3,4,5)P₃, which contains one more phosphate group than PtdIns(4,5)P₂, binding to β_1AR was reduced to a similar level as observed for PI. This demonstrates that binding is selective for the head group of PtdIns(4,5)P₂. We performed similar experiments for NTSR1 and $A_{2A}R$, and in both cases (PtdIns(4,5)P₂) was found to bind with the highest affinity (Extended Data Fig. 3), implying that all three class A GPCRs contain preferential binding sites for PtdIns(4,5)P₂.

We performed coarse-grained molecular dynamics (CGMD) simulations (Extended Data Fig. 4) to characterize the molecular nature of GPCR–PtdIns(4,5)P₂ interactions in a phospholipid bilayer environment⁹. PtdIns(4,5)P₂ molecules bound at the interface formed by the cytoplasmic loops linking transmembrane helix (TM)1, TM2, TM4 and TM7 of NTSR1; this binding was mediated via interactions between the phosphorylated inositol head group and basic protein side chains (Fig. 1c and Extended Data Fig. 4a). Simulation of NTSR1–PS interactions indicated that these were lower-intensity, diffuse interactions that did not compete with PtdIns(4,5)P₂ (Extended Data Fig. 4c). Similar interactions were seen with $\beta_1 AR$, which also exhibited the capacity to interact with PtdIns(4,5)P₂ via the positively charged intracellular surfaces of TM5, TM6 and TM7 (Extended Data Fig. 4b). A more extensive comparison of simulations for nine class A GPCRs (Extended Data

¹Chemical Research Laboratory, University of Oxford, Oxford, UK. ²OMass Technologies, Kidlington, UK. ³Department of Biochemistry, University of Oxford, Oxford, UK. ⁴Biochemisches Institut, Universität Zürich, Zurich, Switzerland. ⁵MRC Laboratory of Molecular Biology, Cambridge, UK. ⁶Present address: Warwick Integrative Synthetic Biology Centre, School of Life Sciences, The University of Warwick, Coventry, UK. *e-mail: mark.sansom@bioch.ox.ac.uk; carol.robinson@chem.ox.ac.uk

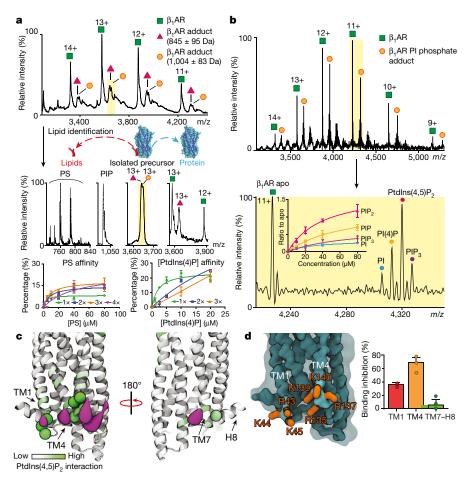


Fig. 1 | Identification of endogenous lipids, preferential binding of PI(4,5)P₂, molecular dynamics simulation and site-directed mutagenesis define intracellular PtdIns(4,5)P₂-binding hotspots. a, Mass spectrum of β_1AR (agonist free, green; charge state is shown) and β_1AR adducts (red, orange). Peaks (highlighted yellow) are selected in the quadrupole and analysed by tandem mass spectrometry. Phosphatidylserine (PS) and PtdIns(4)P (PIP) were identified in the resulting mass spectra. Binding curves plotted against lipid concentration confirm preferential binding of PtdIns(4)P over PS. b, Mass spectra of β_1AR following incubation with an equimolar solution containing PI, PtdIns(4)P, PtdIns(4,5)P₂ and PtdIns(3,4,5)P₃. Binding curves confirm favourable binding of PtdIns(4,5)P₂. c, CGMD simulation for

NTSR1(TM86V- Δ IC3B) embedded in a lipid bilayer containing mixed PC and PtdIns(4,5)P₂. Green spheres represent basic residues with high levels of interaction with lipids; purple surfaces represent regions with high density of occupation by PtdIns(4,5)P₂ (0.6-nm distance cut-off based on the radial distribution of coarse-grained particles). **d**, Left, highlighted residues are mutated in NTSR1(TM86V- Δ IC3B): TM1 (R43G, K44G and K45G; red), TM4 (R135I, R137T, K139L and K140L; orange) and TM7–H8 (R311N; green). Right, inhibition of PtdIns(4,5)P₂ binding. Data are mean \pm s.d. from three independent experiments. Results indicate that mutations on the TM4 interface have a greater effect than those on the TM1 and TM7–H8 interfaces. Binding curves in **a** and **b** are plotted as mean \pm s.d. of three replicates from one experiment.

Fig. 4d) showed that this pattern of interactions with $PtdIns(4,5)P_2$ at the intracellular ends of transmembrane helices is conserved, suggesting that it is structurally and/or functionally significant.

To locate preferential binding sites for PtdIns(4,5)P₂, we performed site-directed mutagenesis on NTSR1, mutating residues that we identified as forming contacts with PtdIns(4,5)P₂ (Fig. 1d) to residues that retain the expression and folded state of the receptor¹⁰. We developed a mass-spectrometry-based strategy to analyse the effect of these mutations on PtdIns(4,5)P₂ binding (Extended Data Fig. 5a). Mutating selected Lys or Arg residues to residues of lower mass decreased the molecular weight of the receptor in comparison to the unmodified parental receptor. When incubated with PtdIns(4,5)P₂, an equimolar solution of mutant and unmodified receptor is presented with an identical lipid environment and can be resolved by mass spectrometry. Attenuation of PtdIns(4,5)P₂ binding was observed in TM1 (35 \pm 0.03%) and TM4 (70 \pm 0.13%) (Fig. 1d and Extended Data Fig. 5b), implying that the cytoplasmic surfaces of these receptors contain hotspots for PtdIns(4,5)P₂ binding.

On the basis of the location of these sites on the cytoplasmic surface, we hypothesized that PtdIns(4,5)P₂ binding influences downstream G-protein coupling. To investigate this, we developed a

mass-spectrometry-based assay in which the pentameric complex of $A_{2A}R$ ($A_{2A}R$ -mini- $G\alpha_s\beta\gamma$ -Nb35; Nb35 is a stabilizing nanobody)^{11,12} was preserved in vacuum. The heteropentamer separated into several subcomplexes following collision-induced dissociation, and PS and PI were observed to be directly bound to $A_{2A}R$ at higher abundance than they were before G-protein coupling (Fig. 2a and Extended Data Fig. 3d). We reasoned that in receptor– $G\alpha\beta\gamma$ complexes, these lipids may have a stabilizing role, thereby, in turn, increasing signalling. To investigate these effects, we measured the GTPase activity of $G\alpha_i\beta\gamma$ when coupled to active NTSR1 (bound to neurotensin₈₋₁₃) in the presence or absence of PtdIns(4,5)P₂. We found that GTP hydrolysis was enhanced to 1.3-fold in the presence of PtdIns(4,5)P₂. Therefore, PtdIns(4,5)P₂ enhances both G-protein coupling and GTPase activity (Fig. 2b).

Because of the instability of the trimeric G-protein complex, it is not possible to explore the effects of lipids on coupling in an unbiased way. We therefore investigated receptor complexes formed with engineered mini-G subunits that recapitulate the increase in agonist affinity observed upon coupling with the native heterotrimeric G protein (Fig. 2c). We recorded mass spectra of thermostabilized β_1AR in complex with mini- G_s . We found increased association of lipids when

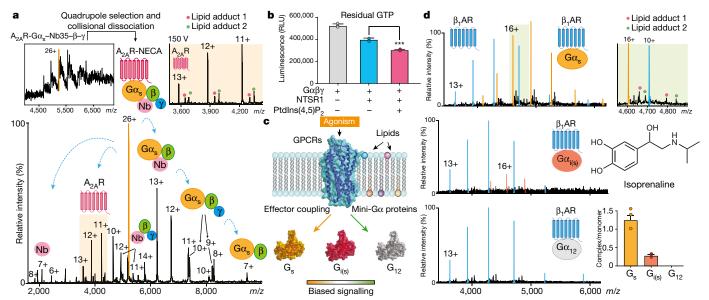


Fig. 2 | Selectivity of G-protein coupling and the presence of endogenous lipids on coupled receptors. a, A representative mass spectrum of $A_{2A}R$ receptor coupled to a trimeric G-protein complex in complex with the stabilizing nanobody Nb35 (top left) from three independent experiments. Isolating and subjecting charge state 26+ (orange peak) to collision-induced dissociation results in dissociation into subcomplexes (bottom) and the receptor with lipid adducts (top right). b, GTPase assays indicate an increase of GTP hydrolysis by active NTSR1 coupled to trimeric $G\alpha_i\beta\gamma$ in the presence of PtdIns(4,5)P2. ****P<0.001;

Student's *t*-test comparing the effect of PtdIns(4,5)P₂ (one variable) on receptor-induced GTPase activation. Bars show mean \pm s.d., points show data from three independent experiments. **c**, Schematic representation of the influence of lipids and agonists on the binding of mini-G proteins. **d**, Mass spectra of isoprenaline-bound β_1AR with three different mini-G subunits (mini-G_s, mini-G_{i(s)} and min_i-G₁₂). Enhanced coupling and lipid adducts are observed in the presence of G_s. (top right) In bottom right, bars show mean \pm s.d., points show data from three independent experiments.

 β_1AR was in a complex with mini- G_s (Fig. 2d). The stability of the receptor–mini- G_s complex allowed us to investigate the selectivity towards different subtypes of $G\alpha$ subunits $(G_s,G\alpha_{i/o}$ and $G\alpha_{12/13}).$ We investigated the coupling of agonist-bound β_1AR to mini- $G_{i(s)}$, which was engineered from mini- G_s by introducing nine mutations on the $\alpha 5$ helix to the corresponding residues on $G\alpha_i$. We performed a similar experiment with the analogous mutant of $G\alpha_{12}$, in which we transferred the mutations from mini- G_s to $G_{12}{}^2$. In comparison to mini- G_s , there was a reduced degree of coupling with mini- $G_{i(s)}$ and virtually no coupling with mini- G_{12} (Fig. 2d).

To investigate the effect of PtdIns(4,5)P₂ on GPCR–mini- G_s interactions, we incubated agonist-bound β_1AR with mini- G_s in the presence of lipid and compared the mass spectrometry peaks corresponding to the lipid-bound protein. Although the complex can form in the absence of lipids, or with only one bound PtdIns(4,5)P₂, complex formation is markedly enhanced (2.7- or 4.5-fold compared to the receptor without lipid, respectively) in the presence of two or three PtdIns(4,5)P₂ molecules (Fig. 3a, g). We observed a similar effect in a time-course experiment in which coupling of mini- G_s to β_1AR increased by $21\pm6\%$ when two PtdIns(4,5)P₂ molecules were bound and by a further $12\pm5\%$ when three PtdIns(4,5)P₂ molecules were bound (Extended Data Fig. 6a).

We examined the effect of PS, an anionic lipid that was endogenously bound to β_1AR (Fig. 1a), on coupling of mini- G_s . We performed analogous experiments using a threefold higher concentration of PS than that used in the experiments with PtdIns(4,5)P $_2$ to reflect the reduced affinity of β_1AR for PS (Fig. 3b and Extended Data Fig. 2). Mass spectra showed only a slight increase in the extent of mini- G_s coupling as a function of PS binding. This reduced effect in comparison to PtdIns(4,5)P $_2$ suggests that the electrostatic interactions of the polyanionic lipid headgroups in PtdIns(4,5)P $_2$, which have multiple basic sidechains, are necessary for receptor coupling (as observed for Kir channels, for example 13), and that such interactions do not occur with PS.

These data indicate that additional $PtdIns(4,5)P_2$, but not PS, stabilize the complex once receptor coupling has occurred. Therefore, we used

potential of mean force (PMF) calculations ¹⁴ to explore the effect of PtdIns(4,5)P₂ binding on the free-energy landscape of $A_{2A}R$ -mini- G_s interactions ¹⁵. Comparison of PMFs for PtdIns(4,5)P₂-bound versus PS-bound receptor in a lipid bilayer indicates that the interaction of mini- G_s with $A_{2A}R$ is stabilized significantly ($50 \pm 10 \text{ kJ mol}^{-1}$) in the presence of PtdIns(4,5)P₂ compared with PS (Fig. 3c and Extended Data Fig. 6b). The presence of PtdIns(4,5)P₂ at the interface between the receptor and mini- G_s in the PMF calculation implies that PtdIns(4,5)P₂ molecules form bridging interactions to stabilize the complex.

The increase in PtdIns(4,5)P₂ binding to β_1 AR when it is coupled to mini-G_s could be a result of either (i) active conformations of receptors binding more PtdIns(4,5)P₂ than their inactive counterparts, or (ii) positively charged residues in mini-G_s, at the receptor-G protein interface, recruiting additional PtdIns(4,5)P2 molecules following coupling. To investigate the dependence of PtdIns(4,5)P₂ binding on receptor conformation, we incubated PtdIns(4,5)P₂ with β_1AR (co-purified with the agonist isoprenaline) containing an E130W mutation to stabilize ligand-free β_1 AR without affecting G-protein coupling ¹⁶. We observed a 31 \pm 1% increase in PtdIns(4,5)P₂ binding to the β_1 AR– isoprenaline complex versus ligand-free β_1AR (Extended Data Fig. 6c). Whereas in general, transition to active states is thought to involve substantial movements of TM5 and TM6, intracellular loop (ICL)2 was also found to undergo significant changes during activation of the κ -opioid receptor¹⁷. These results are consistent with PtdIns(4,5)P₂ stabilizing active states of receptors via binding hotspots directly on ICL2, and, more generaly, via diffuse intracellular PtdIns(4,5)P₂-binding sites.

To explore the second possibility, in which additional PtdIns(4,5) P_2 -binding sites form following coupling, we carried out CGMD simulations for $A_{2A}R$ -mini- G_s , which is, to our knowledge, the only available structure of a receptor-mini-G complex. In addition to the contacts described above, PtdIns(4,5) P_2 interacted with residues of mini- G_s proximal to the lipid contacts in TM3, TM4 and TM5 of $A_{2A}R$ (Fig. 3e). To investigate the significance of these additional binding sites we used a nanobody (Nb6B9) 18 , in which the lipid-binding residues identified in mini- G_s are absent 12 (Extended Data Fig. 7). Structures of receptors bound to Nb6B9 or to mini- G_s are virtually identical 19

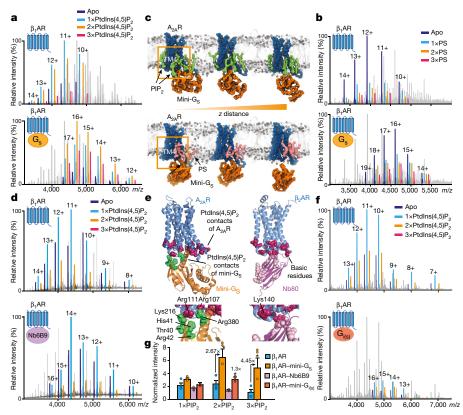


Fig. 3 | The effect of PtdIns(4,5)P₂ on coupling to mini- G_s , and comparison with PS, Nb6B9 and mini- G_i . a, Representative mass spectra of β_1 AR and β_1 AR-mini- G_s (n=3 independent experiments) in the presence of PtdIns(4,5)P₂ and the agonist isoprenaline. Coloured peaks highlight β_1 AR lipid-bound states (top) and β_1 AR-mini- G_s lipid-bound states (bottom). b, Representative mass spectra of β_1 AR and β_1 AR-mini- G_s (n=3 independent experiments) in the presence of PS and the agonist isoprenaline. There is no marked difference in PS binding between β_1 AR and β_1 AR-mini- G_s . c, Snapshots of steered molecular dynamics simulations to separate mini- G_s and A_{2A} R in the presence of PtdIns(4,5) P₂ (green) and PS (pink). Orange outlines highlight the different binding modes of PtdIns(4,5)P₂ and PS to the receptor. The interaction of mini- G_s with A_{2A} R is stabilized by \sim 50 kJ mol⁻¹ in the presence of PtdIns(4,5) P₂ relative to PS (Extended Data Fig. 6b). d, Representative mass spectra

following incubation of β_1AR with PtdIns(4,5)P₂ and isoprenaline in the absence or presence of Nb6B9 (Nb6B9:receptor, 0.3; n=3 independent experiments). **e**, PtdIns(4,5)P₂ contacts on A_{2A}R-mini-G_s are shown on the receptor (purple) and mini-G_s (Thr40, His41, Arg42, Lys216 and Arg380; green), and juxtaposed to basic residues on the β_2AR -Nb80 complex (Nb80, purple). **f**, Representative mass spectra following incubation of β_1AR with PtdIns(4,5)P₂ and isoprenaline in the absence or presence of mini-G_{i(s)} (n=3 independent experiments). No difference was detected between peaks in the presence or absence of PtdIns(4,5)P₂. **g**, Normalized intensity of different lipid-bound states of the apo state of isolated receptor or receptor complexes. *P<0.05; one-way ANOVA with Dunnett's multiple comparison test. Bars show mean \pm s.d., points show data from three independent experiments.

(root mean square displacement (r.m.s.d.) = 0.4–0.6 Å). Comparing PtdIns(4,5)P₂ binding to the receptor and to the receptor–nanobody complex, we found that the degree of PtdIns(4,5)P₂ binding was very similar (Fig. 3d, g). The absence of lipid-binding residues in Nb6B9 (Fig. 3e) explains the insensitivity of the receptor–nanobody complex to PtdIns(4,5)P₂ and implies that PtdIns(4,5)P₂ molecules enhance coupling via interactions that are specific to the receptor and mini-G_s. Lipids such as PS, in which the polyanionic headgroups are absent, would not be able to induce this effect.

To investigate the possibility that residues specific to mini- G_s , that are not present in other G proteins, mediate bridging, we investigated the effects of PtdIns(4,5)P₂ on the coupling of mini- $G_{i(s)}$ to agonist-bound β_1AR . We found that coupling was increased in the presence of PtdIns(4,5)P₂, but to a lesser extent than with mini- G_s (Fig. 3f, g). Given the established role in coupling to receptors of TM5 in $G\alpha_s$ (R380), together with residues identified by molecular dynamics simulation (Fig. 3e), and the fact that these residues are substituted in $G\alpha_i$ (E40, V41, K42, D216 and T380), differences in PtdIns(4,5)P₂-bridging can be attributed to disruption of these PtdIns(4,5)P₂-bridging sites. It therefore follows that PtdIns(4,5)P₂-binding sites on $G\alpha_s$, which are not present on $G\alpha_i$, enable simultaneous binding of the β_1AR to the G protein to which it has highest affinity. Consequently, we propose that PtdIns(4,5)P₂ acts as an allosteric modulator, binding to the

intracellular side of the receptor, stabilizing the active state and enhancing selectivity of G-protein coupling. This coupling is then further stabilized by $PtdIns(4,5)P_2$ molecules bridging between the receptor and the G protein.

More generally, it has been established that the cytoplasmic face of GPCRs undergoes conserved conformational changes to allow coupling of G proteins 20 ; the cytoplasmic ends of TM5 and TM6 move outwards, and TM7 moves slightly inwards. Synthetic molecules that bind at the TM5–TM6–TM7 cytoplasmic interface act as negative allosteric modulators that inhibit the activation of GPCRs by preventing their movement and consequently reducing the affinity of agonists at the orthosteric binding pocket 21,22 . Here we highlight another role of the cytoplasmic interface, which recruits PtdIns(4,5)P₂, thereby stabilizing the active G-protein-bound state of the receptor. Simultaneous binding of the PtdIns(4,5)P₂ head group to both the G α subunit and conserved TM4 residues on a number of class A receptors that are not present on class B receptors suggests the generality of this mechanism for selectively stabilizing active states of class A GPCRs (Extended Data Figs. 4d, 8).

As the local concentration of PtdIns(4,5) P_2 in the membrane has the potential to be modulated by different signalling pathways, such as receptor tyrosine kinases or Ca^{2+} signalling, crosstalk with GPCRs through PtdIns(4,5) P_2 may represent an additional mode of regulation

in the cell²³. Further, the potential to stabilize the active conformation of G-protein-coupled receptors through the binding of potent small molecules that mimic the bridging effects of the PtdIns(4,5)P₂ head group provides a further avenue for stabilizing active states of GPCRs for therapeutic purposes. As PtdIns(4,5)P₂ is able to discriminate between different G-protein subunits, and is likely to also influence binding to β -arrestin, there are potential benefits in developing novel compounds that bind specifically to different G-protein-coupled or β -arrestin-bound states, thereby providing a new perspective for rational design of novel biased allosteric agonists.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0325-6.

Received: 24 July 2017; Accepted: 14 May 2018; Published online 11 July 2018.

- Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* 16, 829–842 (2017).
- Nehmé, R. et al. Mini-G proteins: Novel tools for studying GPCRs in their active conformation. PLoS One 12, e0175642 (2017).
- Zocher, M., Zhang, C., Rasmussen, S. G., Kobilka, B. K. & Muller, D. J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2adrenergic receptor. *Proc. Natl Acad. Sci. USA* 109, E3463–E3472 (2012).
- Dawaliby, R. et al. Allosteric regulation of G protein-coupled receptor activity by phospholipids. Nat. Chem. Biol. 12, 35–39 (2016).
- Gault, J. et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. Nat. Methods 13, 333–336 (2016).
- Yen, H. Y. et al. Ligand binding to a G protein-coupled receptor captured in a mass spectrometer. Sci. Adv. 3, e1701016 (2017).
- Egloff, P. et al. Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli. Proc. Natl Acad. Sci. USA 111, E655–E662 (2014).
- Inagaki, S., Ghirlando, R. & Grisshammer, R. Biophysical characterization of membrane proteins in nanodiscs. *Methods* 59, 287–300 (2013).
- Hedger, G. & Sansom, M. S. P. Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. BBA-Biomembranes 1858, 2390–2400, (2016).
- Schlinkmann, K. M. et al. Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. *Proc. Natl Acad. Sci. USA* 109, 9810–9815 (2012).
- Carpenter, B., Nehmé, R., Warne, T., Leslie, A. G. W. & Tate, C. G. Structure of the adenosine A_{2A} receptor bound to an engineered G protein. *Nature* 536, 104–107 (2016).
- Rasmussen, S. G. et al. Structure of a nanobody-stabilized active state of the β₂ adrenoceptor. *Nature* **469**, 175–180 (2011).
- Hansen, S. B., Tao, X. & MacKinnon, R. Structural basis of PIP₂ activation of the classical inward rectifier K⁺ channel Kir2.2. Nature 477, 495–498 (2011).
- Roux, B. The calculation of the potential of mean force using computer simulations. Comput. Phys. Commun. 91, 275–282 (1995).
- Domański, J., Hedger, G., Best, R. B., Stansfeld, P. J. & Sansom, M. S. P. Convergence and sampling in determining free energy landscapes

- for membrane protein association. J. Phys. Chem. B 121, 3364–3375 (2017).
- Roth, C. B., Hanson, M. A. & Stevens, R. C. Stabilization of the human β₂-adrenergic receptor TM4–TM3–TM5 helix interface by mutagenesis of Glu122^{3.41}, a critical residue in GPCR structure. *J. Mol. Biol.* 376, 1305–1319 (2008).
- 17. Che, T. et al. Structure of the nanobody-stabilized active state of the kappa opioid receptor. *Cell* **172**, 55–67 e15, (2018).
- 18. Miller-Gallacher, J. L. et al. The 2.1 Å resolution structure of cyanopindolol-bound β_1 -adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS One* **9**, e92727 (2014).
- 19. Injin, B. & Hee-Jung, C. Structural features of $\beta 2$ adrenergic receptor: crystal structures and beyond. *Mol. Cells* **38**, 105–111 (2015).
- Venkatakrishnan, A. J. et al. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* 536, 484–487 (2016).
- Zheng, Y. et al. Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature* 540, 458–461 (2016).
- Liu, X. et al. Mechanism of intracellular allosteric β₂AR antagonist revealed by X-ray crystal structure. *Nature* 548, 480–484 (2017).
- Gaví, S., Shumay, E., Wang, H. Y. & Malbon, C. C. G-protein-coupled receptors and tyrosine kinases: crossroads in cell signaling and regulation. *Trends Endocrinol. Metab.* 17, 48–54 (2006).

Acknowledgements C.V.R. acknowledges an ERC Advanced Grant ENABLE (641317), an MRC Programme Grant (G1000819) and a Wellcome Trust Investigator Award (104633/Z/14/Z). Further support was provided by the MRC (G.H.), the Royal Society Newton International Fellowship (W.S.), and the BBSRC (M.S.P.S.; BB/L002558/1) and the Wellcome Trust (M.S.P.S.; 092970/Z/10/Z). This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), supported by EPSRC. A.P. was funded by the Schweizerischer Nationalfonds Grant (31003A_153143). C.G.T. acknowledges the MRC (MC_U105197215), an ERC Advanced Grant EMPSI (339995) and funding from Heptares Therapeutics. We thank M. Hillenbrand for providing Sf9 cells.

Reviewer information *Nature* thanks A. Lee, C. Reynolds, J. Whitelegge and the other anonymous reviewer(s) for their contribution to the peer review of this work

Author contributions H.-Y.Y., K.K.H. and I.L. performed all mass spectrometry experiments on the GPCR, mini- G_{S} and nanobody. D.W. performed lipidomics. G.H., M.R.H., W.S., and M.S.P.S. performed molecular dynamics simulations and analyses. P.H. purified the NTSR1 receptor in the apo state. T.W. purified $\beta_{1}AR$, Y.L. purified $A_{2A}R$ and B.C. purified mini- G_{S} . A.P., C.G.T., M.S.P.S. and C.V.R supervised the research and H.-Y.Y. and C.V.R. wrote the paper with contributions from all authors.

Competing interests H.-Y.Y. and I.L. are founders and employees of OMass Technologies. C.V.R is a founder of and consultant for OMass Technologies.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0325-6

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0325-6.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.S.P.S. or C.V.R.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Constructs and proteins. We used expression plasmids for two stabilized variants of rat NTSR1 7,24 . NTSR1 (HTGH4- Δ IC3B) contains the protein sequence from amino acids 50 to 390 with deletion of ICL3 (residues 273-290) and 26 thermostabilizing point mutations. It should be noted that this construct is only 80% identical to the wild-type. NTSR1(HTGH4 43-421) contains the intact protein sequence from residues 43 to 421, with the same stabilizing mutations as NTSR1(HTGH4- Δ IC3B). Purified thermostabilized turkey (*Meleagris gallopavo*) β_1 AR, human wild-type A_{2A}R, engineered Gα_s (mini-G_S) and nanobody Nb6B9 were used for mass spectrometry analysis 11,25,26 . The following point mutations on β_1AR were used throughout: R68S, M90V, F327A, F338M (thermostabilizing); C116L (to increase protein expression); R284K (residue equivalent to β₂AR designed to improve Nb80 binding); C358A (prevention of potential palmitoylation). In order to purify receptor in the unliganded state, a construct with the same thermostabilizing mutations but slightly different lengths of TM1 was introduced with an additional mutation (E130W) to stabilize the receptor. The use of an N-terminal TrxA fusion (C32S and C35S) on the receptor was necessary to confirm formation of a complex on SDS gels. Insect cell lines for receptor overexpression (Sf9 and Tni) were obtained from Invitrogen and Sf9 cells for heterotrimeric G protein production were provided by M. Hillenbrand. All cells were confirmed to be free from mycoplasma contamination.

Protein expression and purification. *Expression and purification of* β_1AR M. gal*lopavo* β₁AR constructs (β118 and β114-E130W) were based on the previously published thermostabilized β₁AR44-m23 construct²⁷ but contained only four (R68S, M90V, F327A, F338M) of the original six thermostabilizing mutations, as the two mutations on TM5 and TM6 (Y227A and A282L) were not included. The omission of these two mutations resulted in constructs that demonstrated coupling to G proteins and to G protein mimetic nanobody Nb80 along with high affinity agonist binding²⁵. The constructs included E. coli Thioredoxin fused to the N terminus of TM1 and the mutations C116L to improve expression and C358A to prevent potential palmitoylation. Both constructs were expressed in Sf9 insect cells using recombinant baculoviruses prepared using the transfer vector pAcGP67B (BD Biosciences) and BacPAK6 linearized baculovirus DNA (Oxford Expression Technologies). The membrane containing the expressed receptor was solubilized and purified in 2% and 0.02% dodecylmaltoside (DDM, Generon), respectively, as described previously^{27–29}. For β 118, the final purification step was competitive elution from an alprenolol sepharose ligand-affinity column in 20mM Tris-HCl, ph7.4, 350 mM NaCl and 0.02% DDM supplemented with 1mM isoprenaline, so that the receptor was prepared with bound agonist ligand. The purified receptor was finally concentrated to 15 mg/ml in the alprenolol sepharose elution buffer.

β114(E130W) contained the mutation E130W, which increased functional expression of $\beta_1 A R^{16}$. This mutation facilitated the preparation of highly purified active receptor without any bound ligand, as the use of a ligand-affinity chromatography step was not necessary to separate non-functional receptor. For β114(E130W), purification was performed in 0.02% DDM by Ni²⁺ affinity chromatography followed by a thrombin (Sigma) protease cleavage step to remove the His tag before further purification by size-exclusion chromatography (SEC) on a Superdex Increase 200 10/300GL column (GE Healthcare) in 20mM Tris-HCl, ph7.4, 100 mM NaCl and 0.02% DDM, with final concentration to 45 mg/ml. Expression and purification of $A_{2A}R$. The human $A_{2A}R$ construct (residues 1–308) was modified with a C-terminal histidine tag (His10) preceded by a TEV protease cleavage site, and by the mutation N154A to prevent N-linked glycosylation. The A_{2A}R was expressed in Tni insect cells using the baculovirus system. Cell membranes were prepared and solubilised with 2% lauryl maltose neopentyl glycol (LMNG, Anatrace) and the receptor was purified by Ni²⁺ affinity chromatography and SEC, using a Superdex Increase 200 10/300GL column (GE) run in 20 mM HEPES pH 7.5, 100 mM NaCl, 10% (v/v) glycerol, 0.01% (w/v) LMNG and concentrated to 10 mg/ml. Purification was as described previously¹¹, with the exception that the receptor was purified without addition of ligand.

Expression and purification of mini- G_s mini- G_i and mini- G_{12} . The engineered minimal G proteins, mini- G_s construct $R414^{25}$, mini- G_i construct and mini- G_{12} construct 8^2 were expressed in E. coli and purified by Ni^{2+} affinity chromatography, followed by cleavage of the histidine tag using TEV protease and negative purification on Ni^{2+} -NTA to remove TEV and undigested mini-G protein, and finally SEC to remove aggregated protein as described elsewhere 25,30 , with final concentration up to 100 mg/ml in 10 mM HEPES, pH 7.5, 100 mM NaCl, 10% v/v glycerol, 1 mM MgCl₂, 1 μ M GDP and 0.1 mM TCEP.

Expression and purification of nanobody Nb6B9. A synthetic gene (Integrated DNA Technologies) for Nb6B9^{12,31} was cloned into the plasmid pET-26b(+) (Novagen) with a N-terminal His₆ tag followed by a thrombin protease cleavage site. Expression was in *E. coli* strain BL21(DE3)RIL (Agilent Technologies) and purification from the periplasmic fraction was by Ni²⁺ affinity chromatography, but with the use of a thrombin (Sigma) protease cleavage step to remove the His tag before concentration to 40 mg/ml.

Preparation of receptor-G-protein complexes. Several receptor-G-protein complexes were prepared for mass spectrometry analysis. A 2AR-mini-Gs $\beta\gamma$ was prepared by incubating and co-purifying A_{2A}R, containing a TrxA fusion at the N-terminal, with N-ethyl-carboxamidoadenosine (NECA). The complex with trimeric G proteins complex consisted of mini- G_s , $G\beta$, $G\gamma$ and nanobody Nb35 with receptor:G proteins:Nb35 at a 1:2:4 molar ratio to stabilize the complex. The complex was further purified by gel-filtration chromatography after overnight incubation. $\beta_1 AR$ miniG was prepared by incubating β_1AR co-purified with isoprenaline and the different mini-G proteins (mini-G_s, mini-G_{i(s)} and mini-G₁₂) at 1:1.2 molar ratio. The incubation time was varied to capture the equilibrium of complex formation. Purification of heterotrimeric G protein. Baculovirus encoding the desired subunits $(\alpha_{i1}\beta_1\gamma_1)$ was used to express the heterotrimeric G protein in Sf9 cells as previously described³². Cells from a 1-l expression culture were resuspended and lysed in lysis buffer (10 mM HEPES pH 7, 20 mM KCl, 10 mM MgCl₂, $10 \,\mu\text{M}$ GDP, $2 \,\text{mM}$ β-mercaptoethanol, and cOmplete protease inhibitor (Roche)). The membranes were pelleted by ultracentrifugation at 108,000g for 35 min and solubilized in solubilisation buffer (50 mM HEPES pH 7, 150 mM NaCl, 10 mM MgCl $_2$, 10 μ M GDP, $2~\text{mM}~\beta\text{-mercaptoethanol},\,1\%~\text{decyl-}\beta\text{-d-maltopyranoside}$ (DM) (w/v), 10% (v/v) glycerol, and cOmplete protease inhibitor (Roche)) for 3 h. The supernatant was collected after centrifugation at 108,000g for 35 min and incubated with 1.2 ml TALON beads (GE Healthcare) overnight. The beads were collected and washed with ten column volumes wash buffer (30 mM HEPES pH 7, 300 mM NaCl, 10 mM MgCl₂, 25 mM imidazole pH 8, 10 μ M GDP, 2 mM β -mercaptoethanol, 10% (v/v) glycerol, and 0.5% (w/v) DM), followed by another twenty columnvolume wash of wash buffer containing 40 mM imidazole (pH 8.0), and were eluted with five column volumes elution buffer (30 mM HEPES pH 7, 150 mM NaCl, 1 mM MgCl₂, 300 mM imidazole pH 8, 10 μ M GDP, 2 mM β -mercaptoethanol, 10% (v/)v) glycerol, and 0.5% (w/v) DM). The protein was further purified by a Superdex 200 Increase PC 3.2/300 column (GE Healthcare) and the protein tag was removed by incubation with human rhinovirus 3C protease (produced in house) overnight. Following buffer exchange to storage buffer (20 mM HEPES pH 7, 100 mM NaCl, 0.1 mM MgCl₂, 4 mM β -mercaptoethanol, 10% (v/)v) glycerol, and 0.5% (w/v) DM) and reverse immoblized metal affinity chromatography (IMAC) by Ni-NTA superflow beads (GE Healthcare), G-protein complex was concentrated to at least 2 mg/ml for experimental use.

NTSR1 expression BL21 E. coli cells were transformed with the expression plasmid encoding NTSR1(HTGH4- Δ IC3B) and grown overnight at 37 °C in 20 ml 2YT medium supplemented with 1% (w/v) glucose and 100 µg/ml ampicillin. Two flasks, each containing each 112YT medium, 0.5% (w/v) glucose, and 100 µg/ml ampicillin were inoculated with 10 ml pre-culture and grown to an $A_{600\,\mathrm{nm}}$ of 0.5 with shaking at 37 °C. Receptor expression was induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and cells were cultivated at 28 °C overnight. Cells were harvested after overnight expression and E. coli cell pellets were resuspended in 100 ml solubilisation buffer, containing 100 mM HEPES pH 8.0, 20% (v/v) glycerol and 400 mM NaCl. Resuspended cells were frozen in liquid nitrogen and stored at -80 °C.

Apo NTSR1 purification. The cell pellet was thawed at room temperature. All following steps were carried out at 4 °C. MgCl₂ (5 mM), 2 mg DNase I, 200 mg lysozyme, and 20 ml detergent mixture (0.2% (w/v) cholesteryl hemisuccinate Tris salt (CHS) and 2% (w/v) dodecyl-β-D-maltopyranoside (DDM)) were added to the thawed cell pellet. The mixture was incubated for 1 h, followed by cell lysis via mild sonication for 30 min in an ice-water bath. After cell lysis, 0.4 ml 5 M imidazole was added and the mixture was incubated for another 30 min. The suspension was centrifuged for 30 min at 28,000g. The supernatant was mixed with 5 ml TALON resin (Clontech), which had been pre-equilibrated with IMAC binding buffer (25 mM HEPES pH 8.0, 10% (v/v) glycerol, 600 mM NaCl, 0.1% (w/v) DDM and 20 mM imidazole) and incubated overnight on a rolling device. The mixture was loaded into a PD10 column (GE Healthcare) and was washed with 50 ml IMAC binding buffer. Elution of bound protein was performed with 15 ml IMAC elution buffer containing 25 mM Hepes pH 8.0, 10% (v/v) glycerol, 150 mM NaCl, 0.1% (w/v) DDM and 250 mM imidazole. Eluted receptor was concentrated in an Amicon-15 Ultra concentrator with a 100 kDa cut-off (Millipore) to a final volume of less than 2.5 ml. Concentrated receptor sample was loaded on a Sephadex G-25 desalting column (GE Healthcare), pre-equilibrated with 25 mM Hepes pH 8.0, 10% (v/v) glycerol, 150 mM NaCl, 0.1% (w/v) DDM to remove remaining imidazole. Desalted receptor was incubated with $300\,\mu l$ 1.6 mg/ml HRV 3C protease for 1 h at 4 °C, followed by addition of 150 μl 10% (w/v) LMNG and incubation for 1 h at 4°C. The cleaved protein was diluted threefold with reverse IMAC buffer (10 mM HEPES pH 8.0, 10% (v/v) glycerol, 150 mM NaCl, and 0.01% (w/v) LMNG) and was loaded onto a PD10 column containing 5 ml Ni-NTA beads pre-equilibrated with reverse IMAC buffer. The flow through was collected in an Amicon-15 Ultra concentrator with a 50-kDa cut-off and the resin was further washed with 10 ml buffer. Receptor was concentrated to a final volume of less than 1 ml and was subjected to preparative

SEC using a Superdex 200 10/300 GL column (GE Healthcare), which had been pre-equilibrated with 10 mM HEPES pH 8, 150 mM NaCl, and 0.01% (w/v) LMNG. Peak fractions corresponding to NTSR1(HTGH4- Δ IC3B) were pooled (final volume 3-4 ml) and concentrated in an Amicon-4 Ultra-concentrator with a 50-kDa cut-off to a final protein concentration of approximately 50 μ M. Purified and concentrated NTSR1-H4 was mixed with 10 mM HEPES pH 8, 150 mM NaCl, 0.01% (w/v) LMNG, and 50% (v/v) glycerol to yield a final glycerol concentration of 25%. Aliquots were frozen in liquid nitrogen and stored at -80 °C for later use. Preparation of phospholipids and titration experiment. Phospholipids were purchased from Avanti (Avanti Polar Lipids) and prepared as 3 mM stock solutions in 200 mM ammonium acetate buffer pH 7.5 containing the detergent-mixed micelle preparation, containing DDM and foscholine as previously described³³. Phosphate analysis was performed to determine the concentration of phospholipids in solution³⁴. For the titration experiment, 5 μM buffer-exchanged receptors in 200 mM ammonium acetate buffer pH 7.5 containing the detergent mixtures (DDM, LMNG, and foscholine for NTSR1; DDM and foscholine for β_1AR and A2AR) were mixed with lipids at various concentration points followed by equilibration at $4\,^{\circ}\text{C}$ for 5 min, by which time lipid binding had stabilized according to our time course measurements. Following mass spectrometry analysis, UniDec (Universal Deconvolution) software was used to quantify the relative abundance of each lipid-bound state³⁵, and statistical analysis was performed using GraphPad Prism, assuming a one-site total binding model.

Lipidomics analysis. Co-purified lipids from recombinant GPCRs were extracted by chloroform-methanol (2:1, v/v) and lyophilized and re-dissolved in 60% acetonitrile (ACN). For LC-MS/MS analysis, the extracted lipids were separated on a C18 column (Acclaim PepMap 100, C18, 75 mm × 15 cm; Thermo Scientific) using a Dionex UltiMate 3000 RSLC nano LC System. The buffers and gradient are adapted from a previous protocol³⁶. In brief, the lipids were separated using a binary buffer system at 40 °C using a gradient of 32-99% buffer B at a flow rate of 300 nl/min over 30 min. (Buffer A: (acetonitrile: H₂O (60:40), 10 mM ammonium formate, 0.1% formic acid) and buffer B (propan-2-ol:acetonitrile (90:10), 10~mMammonium formate, 0,1% formic acid)). The column eluent was delivered via a dynamic nanospray source to a hybrid LTQ Orbitrap mass spectrometer (Thermo Scientific). Typical mass spectrometry conditions were: spray voltage (1.8 kV) and capillary temperature (175 °C). The LTQ-Orbitrap XL was operated in negative ion mode using data-dependent acquisition with one MS scan followed by five MS/MS scans³⁷. Survey full-scan mass spectra were acquired in the orbitrap (m/z 350-2,000) with a resolution of 60,000. CID fragmentation in the linear ion trap was performed for the five most intense ions at an automatic gain control target of 30,000 and a normalized collision energy of 38% at an activation of q = 0.25 and an activation time of 30 ms.

GTPase assay. The GTPase activity of trimeric $G\alpha_i\beta\gamma$ was measured with the GTPase-Glo assay (Promega). The assay was performed in white 384-well plates (Corning) using purified trimeric G proteins diluted into a GTPase buffer (10 mM HEPES pH 7, 50 mM NaCl, 0.05 mM MgCl₂, 2 mM β -mercaptoethanol, 1mM DTT, 5% (v/v) glycerol, and 0.25% (w/v) DM) at a finial concentration 2.5 μ M in the presence of 5 μ M GTP. The luminescent signal was measured after incubation at room temperature (1 h) following the manufacturer's protocol to indicate the level of residual GTP. To analyse the impact of PtdIns(4,5)P₂ we used NTSR1(HTGH4- Δ IC3B) co-purified with recombinant neurotensin₈₋₁₃ following the method described previously³⁸. The receptor was pre-incubated with detergent-solubilised PtdIns(4,5)P₂ at 1:3 molar ratio (receptor:lipid) in the protein buffer (10 mM HEPES pH 8, 150 mM NaCl, 0.01% (w/v) LMNG) containing 100 nM neurotensin₈₋₁₃ for 15 min on ice. The activated receptor was then added to the reaction mixture containing trimeric G proteins under the condition described above.

Native mass spectrometry of GPCRs. Purified GPCRs were buffer exchanged into 200 mM ammonium acetate buffer pH 7.5 containing the mixed micelle preparation optimized for GPCR analysis as described previously 6 . The concentration of DDM, foscholine and CHS required to form a mixed micelle range from 0.006–0.02%, 0–0.002%, and 0.001–0.01%, respectively, and are optimized for each receptor preparation. The samples were immediately introduced into a modified Q-Exactive mass spectrometer (Thermo), as described previously 5 . Ions were transferred into the higher-energy collisional dissociation (HCD) cell following a gentle voltage gradient (injection flatapole, inter-flatapole lens, bent flatapole, transfer multipole: 7.9, 6.94, 5.9, 4 V, respectively). An optimized acceleration voltage (100–130 V) was then applied to the HCD cell to remove the detergent micelle from the protein ions. Backing pressure was maintained at $\sim 1.00 \times 10^{-9}$ mbar and data was analysed using Xcalibur 2.2 SP1.48.

The bound-lipid identification experiments were performed with a modified Synapt G2 mass spectrometer (Waters) equipped with a Z-spray source 33,39 . The typical instrumental setting was source pressure (4.5–5.0 mbar), capillary voltage (1.2–1.5 kV) and cone voltage (100–200 V). An extraction voltage of 1–5 V was applied and 80–150 V was used as the collision voltage with argon as the collision gas

at a pressure of 0.2–0.3 MPa. To strip the detergent from protein ions in the source region, instrument values were optimized to capillary voltage $(1.5~{\rm kV})$, cone voltage $(200~{\rm V})$ and extraction voltage $(3~{\rm V})$. A collision voltage ramp (from 20–100 V) was applied to dissociate protein–lipid complexes after quadrupole selection.

Identification of preferential PtdIns(4,5)P₂-binding sites on NTSR1. Unmodified NTSR1 and NTSR1 variants were pre-incubated at 1:1 molar ratio to produce a total protein concentration of 12 mM in protein buffer (10 mM HEPES pH 8, 150 mM NaCl, 0.01% (w/v) LMNG and 25% (v/v) glycerol). Detergent solubilised PI(4,5)P₂ was then added to the protein mixture at a final molar ratio of 1.25:1 lipid:receptor. The reaction mixture was incubated at 4° C for 5 min and analysed by mass spectrometry after buffer exchanging to 200 mM ammonium acetate buffer containing the mix of detergents of DDM, LMNG and foscholine as described previously⁶.

The ratio of PtdIns(4,5) P_2 binding to the receptor was calculated by normalizing the intensity of the receptor in PtdIns(4,5) P_2 -bound states to the unbound state using UniDec software. The results were evaluated by comparing the ratio of PtdIns(4,5) P_2 binding between mutants and the unmodified receptor and plotted as a bar chart using GraphPad Prism.

Mini- G_S and Nb6B9 coupling to β_1AR . Effector coupling to β_1AR was analysed using a modified Q-Exactive mass spectrometer after incubating purified β_1AR with mini- G_S -Nb6B9 at 1:1.2 molar ratio at 4°C in protein buffer (20mM Tris-HCl, ph7.4, 350 mM NaCl and 0.02% DDM). The relative percentage of effector coupling was quantified by UniDec software. A time course was performed with aliquots sampled after 2, 10, 30, and 60 min to monitor the formation of the mini- G_S -receptor complex. To investigate the effect of PtdIns(4,5)P₂ on coupling, β_1AR was pre-incubated with detergent-solubilised PtdIns(4,5)P₂ at 1:1 molar ratio for 5 min at 4°C to equilibrate before mixing with mini- G_S or Nb6B9 at 1.2 or 0.3 molar ratio to receptor, respectively. For the analogous PS binding experiment we pre-incubated β_1AR with a threefold higher concentration of detergent solubilised PS than PtdIns(4,5)P2 (PS: β_1AR , 3:1 molar ratio) for 5 min at 4°C to equilibrate before mixing with mini- G_S .

Modelling and simulation system setup. Simulations were performed using the GROMACS v.4.6.3 simulation package. Initial protein coordinates were obtained using PDB ID 4BUO (NTSR1) and PDB ID 2Y03 (β_1AR), with missing atoms added using MODELLER⁴⁰. In the case of β_1AR , a model was also constructed in which S68 in the thermostabilized structure 2Y03 was back-mutated to R68 to reconstruct available basic residues in the wild-type receptor using the mutagenesis tool implemented in PyMOL v.1.3r1. Side-chain ionisation states were modelled using pdb2gmx⁴¹. The N and C termini were treated with neutral charge. Each protein structure was then energy minimized using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained representation using the MARTINI 2.2 force field⁴². The energy minimized coarse-grained structure was centred in a periodic simulation box with dimensions $11 \times 11 \times 12 \text{ nm}^3$. POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) molecules were randomly placed around the protein and the system was solvated and neutralised to a concentration of 0.15 M NaCl. An initial 50 ns of coarsegrained simulation was applied to permit the self-assembly of a POPC lipid bilayer around the GPCR. POPC lipids were randomly exchanged⁴³ to create a mixedspecies bilayer of specified composition (Extended Data Table 2). A cut-off distance of 2.5 nm was applied, with only molecules outside this distance being subject to exchange. The exchange protocol was conducted independently for each repeat simulation, such that different random initial configurations of lipids around the protein were generated for each simulation repeat. A summary of simulations performed is provided in Extended Data Table 2.

Simulation details. The MARTINI force field⁴² was used to describe all system components. An ELNEDYN network⁴⁴ was applied to the protein using a force constant of 500 kJ/mol/nm² and a cut off of 1.5 nm. Simulations were performed as an NPT ensemble, with temperature maintained at 310 K using a Berendsen thermostat⁴⁵ using a coupling constant of $\tau_{\rm t}$ = 4 ps, and semi-isotropic pressure controlled at 1 bar using a Berendsen barostat⁴⁵ with a coupling constant of $\tau_{\rm p}$ = 4 ps and a compressibility of 5 × 10⁻⁶ bar¹. Electrostatics were modelled using the reaction field coulomb type⁴⁶, and smoothly shifted between 0 and 1.2 nm. Van der Waals interactions were treated using a shifting function between 0.9 and 1.2 nm. Covalent bonds were constrained to their equilibrium values using the LINCS algorithm⁴⁷. Equations of motion were integrated using the leap-frog algorithm, with a 20-fs time step. All simulations were run in the presence of conventional MARTINI water, and neutralised to a concentration of 0.15 M NaCl.

Analysis of simulation data was conducted using VMD^{48} , PyMOL, tools implemented in GROMACS⁴¹, and in-house protocols. Protein–lipid contact analysis employed a cut-off distance of 0.6 nm, based on radial distribution functions for coarse-grained lipid molecules⁴⁹.

 $\label{eq:A2AR-mini-G3PMF} \textbf{A2AR-mini-G3PMF calculations.} \ PMFs \ for the interaction of mini-G3 with A2AR in a lipid bilayer in the presence and absence of PtdIns(4,5)P2 were calculated using the MARTINI force field 50. To obtain a PtdIns(4,5)P2-bound A2AR-mini-G3 complex, $$$

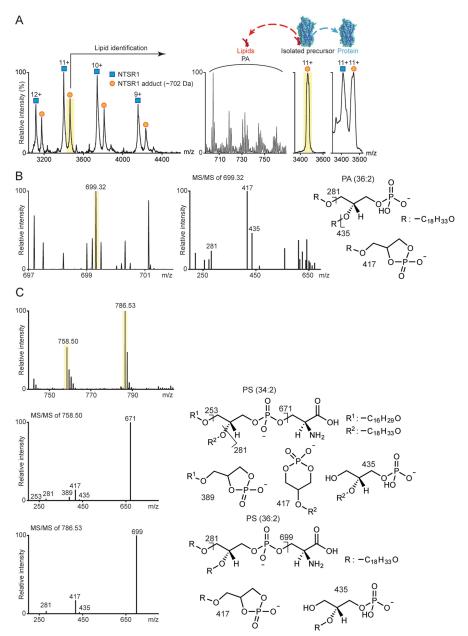
we first ran ten coarse-grained molecular dynamics simulations on receptor embedded in an asymmetric complex membrane, each lasting 8 µs (Extended Data Table 2). The r.m.s.d. to the crystal structure of A_{2A}R-mini-G_S complex (PDB ID 5G53) was calculated for the protein in these ten simulations, and the protein complex with the lowest r.m.s.d. was saved together with the membrane bilayer. The coarse grained mini-G_S was then docked back to the membraneembedded receptor based on the A_{2A}R-mini-G_S crystal structure to generate the starting configuration of a steered molecular dynamics (SMD) simulation. In the SMD, the mini-G_S was pulled away from the receptor along the z axis (normal to the membrane plane) at a rate of 0.05 nm/ns using a force constant of 1000 kJ/mol/nm² while the receptor was restrained in place using a harmonic force of 1000 kJ/mol/nm². The distance between the centre of mass of the receptor and the mini-G_S was defined as the 1D reaction coordinate and the pulling process covered a distance of 3 nm. The initial configurations of the umbrella sampling were extracted from the SMD trajectory spacing 0.05 nm apart along the reaction coordinate. Fifty umbrella sampling windows were generated, and each was subjected to 1-µs molecular dynamics simulation, in which a harmonic restrain of 1000 kJ/mol/nm² was imposed on the distance between the centre of mass of the receptor and the mini-G_S to maintain the separation of the two. The PMF was extracted from the umbrella sampling using the weighted histogram analysis method (WHAM) provided by the GROMACS g_wham tool⁵¹. A Bayesian bootstrap was used to estimate the statistical error of the energy profile. The PMF of the binding process in the absence of PtdIns(4,5)P2 was calculated following the same protocol, with the only change made to the lipid composition of the membrane lower leaflet. PtdIns(4,5)P₂ was taken out from the membrane and instead the concentrations of POPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) and 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) were increased by 2.5% to make up for the vacancy left by the absence of PtdIns(4,5)P₂.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. All relevant data are available from corresponding authors on request.

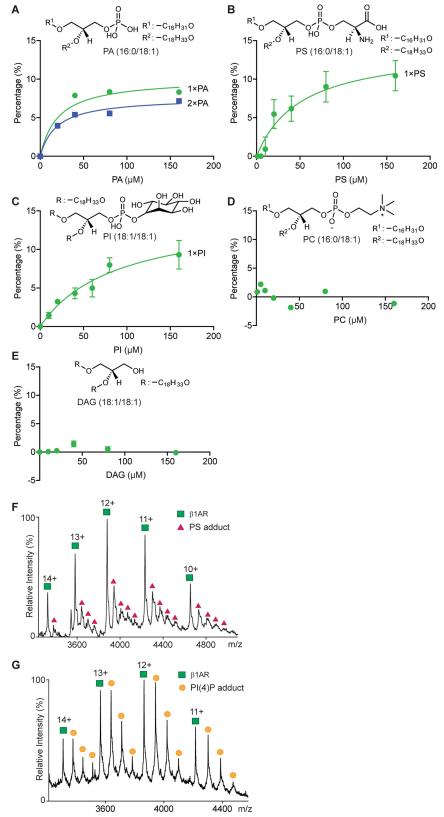
- Scott, D. J., Kummer, L., Egloff, P., Bathgate, R. A. D. & Plückthun, A. Improving the apo-state detergent stability of NTS₁ with CHESS for pharmacological and structural studies. *BBA-Biomembranes* **1838**, 2817–2824, (2014).
- Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate crystallisation of G protein-coupled receptors in their active conformation. *Protein Eng. Des. Sel.* 29, 583–594 (2016).
- 26. Warne, T. et al. Structure of a β_1 -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- 434, 485–491 (2008). 27. Warne, T. et al. The structural basis for agonist and partial agonist action on a \$\beta_1\$-adrenergic receptor. *Nature* **469**, 241–244 (2011).
- Warne, T., Serrano-Vega, M. J., Tate, C. G. & Schertler, G. F. Development and crystallization of a minimal thermostabilised G protein-coupled receptor. *Protein Expr. Purif.* 65, 204–213 (2009).
- Warne, T., Chirnside, J. & Schertler, G. F. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim. Biophys. Acta* 1610, 133–140 (2003).
- Carpenter, B. & Tate, C. G. Expression and Purification of Mini G Proteins from Escherichia coli. Bio Protoc. 7, (2017).
- 31. Ring, A. M. et al. Adrenaline-activated structure of β_2 -adrenoceptor stabilized by an engineered nanobody. *Nature* **502**, 575–579 (2013).

- 32. Hillenbrand, M., Schori, C., Schoppe, J. & Pluckthun, A. Comprehensive analysis of heterotrimeric G-protein complex diversity and their interactions with GPCRs in solution. *Proc. Natl Acad. Sci. USA* **112**, E1181–E1190 (2015).
- Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protocols* 8, 639–651 (2013).
- Chen, P. S., Toribara, T. Y. & Warner, H. Microdetermination of phosphorus. Anal. Chem. 28, 1756–1758 (1956).
- Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* 87, 4370–4376 (2015).
- Bird, S. S., Marur, V. R., Sniatynski, M. J., Greenberg, H. K. & Kristal, B. S. Lipidomics profiling by high resolution LC–MS and HCD fragmentation: focus on characterization of mitochondrial cardiolipins and monolysocardiolipins. *Anal. Chem.* 83, 940–949 (2011).
- 37. Bechara, C. et al. A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nat. Chem.* **7**, 255–262 (2015).
- Egloff, P., Deluigi, M., Heine, P., Balada, S. & Pluckthun, A. A cleavable ligand column for the rapid isolation of large quantities of homogeneous and functional neurotensin receptor 1 variants from E. coli. Protein Expr. Purif. 108, 106–114 (2015).
- Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* 74, 1402–1407 (2002).
- Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993).
- Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447 (2008).
- de Jong, D. H. et al. Improved parameters for the Martini coarse-grained protein force field. J. Chem. Theory Comput. 9, 687–697 (2013).
- Koldsø, H., Shorthouse, D., Hélie, J. & Sansom, M. S. P. Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers. *PLOS Comput. Biol.* 10, e1003911 (2014).
- Periole, X., Cavalli, M., Marrink, S. J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* 5, 2531–2543 (2009).
- Berendsen, H. J. C., Postma, J. P. M., van Gúnsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684–3690 (1984).
- Tironi, I. G., Sperb, R., Smith, P. E. & van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* 102, 5451–5459 (1995).
- Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472 (1997)
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38 (1996).
- Hedger, G., Sansom, M. S. P. & Koldsø, H. The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. Sci. Rep. 5, 9198 (2015).
- Hedger, G., Shorthouse, D., Koldso, H. & Sansom, M. S. Free energy landscape of lipid interactions with regulatory binding sites on the transmembrane domain of the EGF receptor. J. Phys. Chem. B 120, 8154–8163 (2016).
- Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—A free weighted histogram analysis implementation including robust error and autocorrelation estimates. J. Chem. Theory Comput. 6, 3713–3720 (2010).
- Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217 (2000).
- Bond, C. S. & Schuttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. D 65, 510–512 (2009).



Extended Data Fig. 1 | Identification of lipids bound to NTSR1(HTGH4- Δ IC3B). a, Endogenous lipids bound to NTSR1(HTGH4- Δ IC3B), isolated from *E. coli*, are identified as PA following m/z selection in the mass spectrometry quadrupole of the NTSR1:lipid 11+ charge state (highlighted yellow) and collisional activation to dissociate PA and its homologues (m/z, 700–760 Da). b, Lipidomics analysis of purified NTSR1 with three technical replicates

reveals peaks at low m/z. MS/MS spectra of the precursor ion (M-H-1) at m/z 699.32 highlighted yellow, leads to definitive fragment ions at m/z 281 and 417 consistent with the structure of PA (36:2). **c**, Analogous lipidomics analysis of purified $\beta_1 AR$ from insect cells with three technical replicates. MS/MS spectra of the two [M-H-1] precursor ions (m/z 758.50 and 786.53) identified the lipids as PS (34:2) and PS (36:2) respectively with diagnostic fragments indicated.

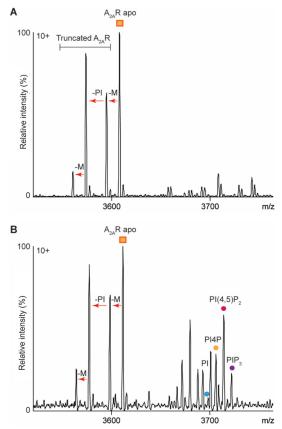


Extended Data Fig. 2 | See next page for caption.

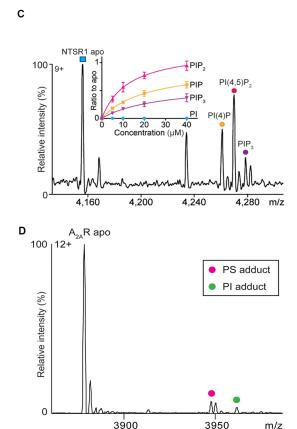


Extended Data Fig. 2 | Lipid-binding preference of NTSR1 and $\beta_1 AR$. a–e, The binding of NTSR1(HTGH4- Δ IC3B), measured by mass spectrometry (n=3 independent experiments), to the phospholipids PA (a), PS (b), PI (c), PC (d) and DAG (e). The measurements were performed at different lipid concentrations (0 to 160 μ M) and the percentages of individual lipid-binding peaks (sum of apo protein and all lipid adducts obtained in the region of the mass spectrum under study) were plotted against lipid concentrations in solution. The lipid-binding curves were deduced from fitting to one-site total binding. Values of s.d. were calculated from three independent replicate experiments at each concentration. The results show that NTSR1 interacts preferentially

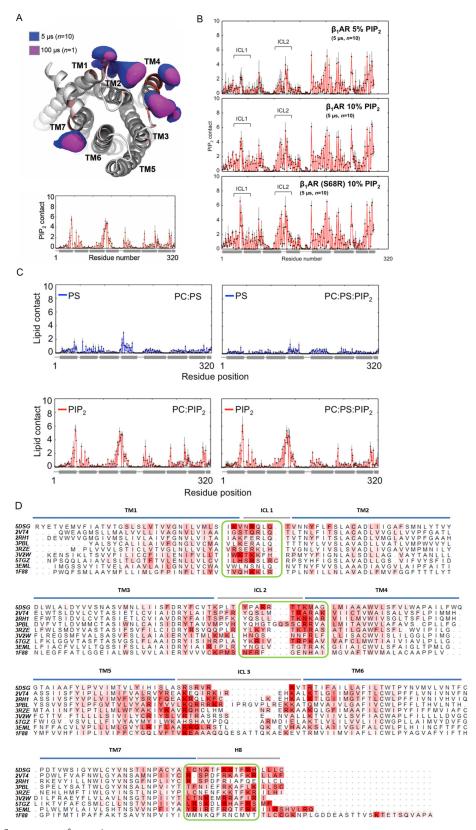
with anionic phospholipids (PA and PS), as no binding was observed for neutral (DAG) and zwitterionic (PC) lipids. **f**, **g**, Exogenous POPS (**f**) and PtdIns(4)P (**g**) were added to β_1AR at different final concentrations (10 μ M is shown here). Spectra were recorded for a range of lipid concentrations from 0 to 80 μ M for PS and 0 to 20 μ M for PtdIns(4)P. Peak intensities of the individual PtdIns(4)P-bound species were measured and plotted against lipid concentration to yield a relative affinity for one PtdIns(4)P binding (1×), two PtdIns(4)P molecules binding (2×) or three PtdIns(4)P molecules binding (3×); only the first PtdIns(4)P molecule binds with high affinity (see Fig. 1a). Data are mean \pm s.d. from three independent experiments.



Extended Data Fig. 3 | Investigation of the phospholipid preferences of $A_{2A}R$ and NTSR1. a, A representative mass spectrum of purified $A_{2A}R$ from three independent experiments revealed truncations of the N-terminal sequence (MPIM). The arrows between species indicate the mass differences corresponding to truncated amino acids (M, PI and M). b, A competitive binding assay (n=3 independent experiments) in which $A_{2A}R$ was incubated with a mixture of lipids (PI, PtdIns(4)P, PI(4,5)P₂, and PtdIns(3,4,5)P₃) before mass spectrometry, indicated that PtdIns(4,5)P₂ binds with a higher affinity than the other phospholipids to $A_{2A}R$. c, The analogous competitive binding assay, in which NTSR1 was incubated with



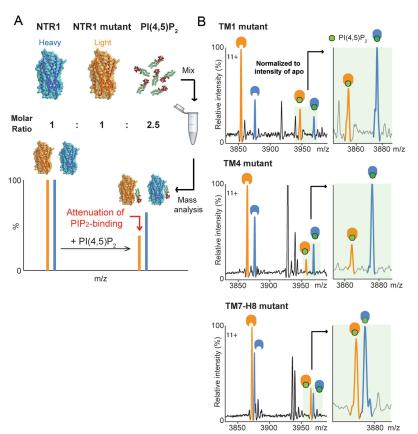
a mixture of lipids (PI, PtdIns(4)P, PI(4,5)P₂ and PtdIns(3,4,5)P₃) before mass spectrometry. Ratio to apo is plotted as a function of concentration and defined as the ratio of the intensity corresponding to individual PI phosphate adducts to the receptor in the apo state (inset). The same data analysis methods are used for Fig. 1b. PtdIns(4,5)P₂ binds with a higher affinity than the other phospholipids to $A_{2A}R$. Data are shown as mean \pm s.d. from three independent replicates. \boldsymbol{d} , A representative mass spectrum of $A_{2A}R$ (n=3 independent experiments) used for preparation of the G-protein complex reveals lower abundance of PS and PI adducts prior to coupling to G proteins.



Extended Data Fig. 4 \mid See next page for caption.

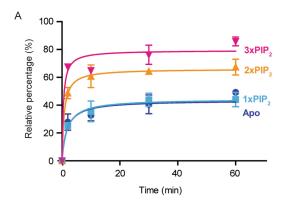
Extended Data Fig. 4 | NTSR1-PtdIns(4,5)P₂ and β₁AR-PtdIns(4,5)P₂ interactions within CGMD simulations, and comparison of PtdIns(4,5) P2 contacts among different GPCRs. a, Volumetric density surfaces showing the average spatial occupancy of PtdIns(4,5)P2 lipids around a crystal structure of NTSR1(TM86V-ΔIC3B) (PDB: 4BUO), which shares a greater sequence identity to the wild-type receptor (91%) than NTSR1(HTGH4-ΔIC3B) (86%), contoured to show the major PtdIns(4,5) P₂-interaction sites. Density surfaces were calculated over 5 µs of CGMD (blue surface, n = 10 independent experiments), and 100 µs of CGMD (magenta, n = 1 experiment). The cytoplasmic side of NTSR1 structure is coloured from white (low PtdIns(4,5)P2 interaction) to red (high PtdIns(4,5)P₂ interaction). Extending a simulation to 100 μs revealed no overall change in the patterns of PtdIns(4,5)P₂ interaction. Less specific, and hence more dynamic, interaction was seen for the acyl chain moieties of PtdIns(4,5)P₂, which yielded more diffuse probability densities. \mathbf{b} , $\beta_1 AR$ -PtdIns(4,5) P_2 interactions within CGMD simulations. Contact patterns are shown for simulations containing 5% PtdIns(4,5)P₂ in the lipid bilayer and thermostable β_1AR (PDB: 2Y03, top), 10% PtdIns(4,5)P₂ and thermostable β_1AR (middle), and 10% PtdIns(4,5)P₂ and β_1AR (S68R) construct (bottom). In each case $PtdIns(4,5)P_2$ contacts were calculated over 5 μ s of CGMD (n = 10 independent experiments; error bars, s.d.), with each repeat simulation initiated from different random system

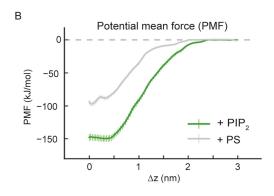
configurations. c, PS and PtdIns(4,5)P2 contacts with NTSR1 as a function of residue position, for PC:PS membranes (top left), PC:PS:PtdIns(4,5) P₂ membranes (top right), PC:PtdIns(4,5)P₂ membrane (bottom left) and PC:PS:PtdIns(4,5)P₂ (bottom right). The position of helices is denoted by horizontal grey bars. Lipid contact is calculated as the mean number of contacts between each residue and a given lipid species per frame, using a 6 Å distance cut-off. n = 3; error bars, s.d.. **d**, PtdIns(4,5)P₂ contacts seen in CGMD simulations for nine class A GPCRs: histamine H1 receptor, PDB 3RZE; β_1 adrenergic receptor, 2VT4; β_2 adrenergic receptor, 2RH1; CB1 cannabinoid receptor, 5TGZ; M4 muscarinic acetylcholine receptor, 5DSG; adenosine A_{2A} receptor, 3EML; dopamine D3 receptor, 3PBL; sphingosine 1-phosphate receptor, 3V2W; rhodopsin, 1F88. GPCR sequences are shown, with TM helices, intracellular loops (ICL) and H8 helices indicated by horizontal bars, and with amino acids coloured according to the mean number of contacts per simulation frame with the PtdIns(4,5)P₂ molecules. Green boxes correspond to the high frequency of PtdIns(4,5)P₂ interactions discussed in the main text for the TM1, TM4, and TM7/H8 motifs of NTSR1. Contacts were computed over 1 μs CGMD simulations (n = 3 independent experiments) for each GPCR, using a 6 Å cut-off. Sequences were aligned using T-Coffee⁵² and mapping of proteinlipid contact data onto the sequence alignment used ALINE⁵³

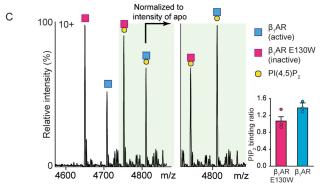


Extended Data Fig. 5 | Site-directed mutagenesis attenuates PtdIns(4,5) P_2 binding to NSTR1. a, Schematic representation of the experimental protocol designed to combine mass spectrometry with mutagenesis to produce mutants of lower molecular mass than wild type, which, when incubated with PtdIns(4,5) P_2 , yield a direct readout of the effect of mutations in specific regions. b, PtdIns(4,5) P_2 binding of NTSR1 mutants on residues that exhibit the highest frequency of PtdIns(4,5) P_2 interaction in molecular dynamics simulation. Mutation of NTSR1(HTGH4- Δ IC3B) residues on TM1 (R46G, K47G and K48G (R43G, K44G and K45G in NTSR1(TM86- Δ IC3B); R91G, K92G, K93G in wild type)), TM4

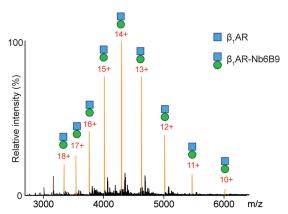
(R138I, R140T, K142L and K143L (R135I, R137T, K139L and K140L in NTSR1(TM86- Δ IC3B); R183I, R185T, K187L and K188L in wild type)) and TM7-H8 (R316N (R311N in NTSR1(TM86- Δ IC3B); R377N in wild type)) attenuate PtdIns(4,5)P $_2$ binding, and indicate that the TM4 interface is a preferential binding site over TM1 and TM7-H8 interfaces. Selection of residues for mutations was guided by molecular dynamics (Extended Data Fig. 4) and previous studies in which binding of a fluorescently labelled agonist, BODIPY neurotensin, to NTSR1, was screened and used to monitor efficient production, insertion, and folding 10 .



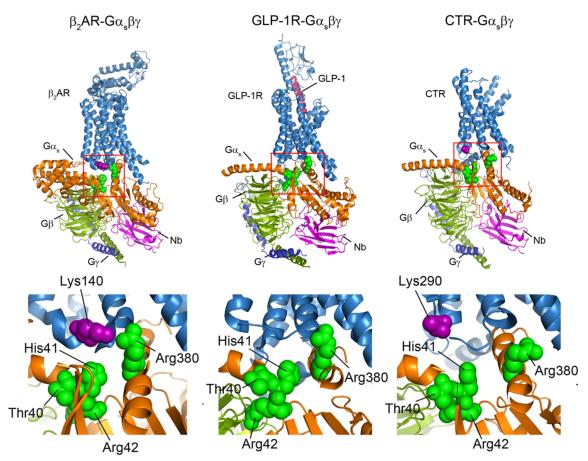




Extended Data Fig. 6 | PtdIns(4,5)P₂ binds preferentially to β_1AR in an active state and stabilizes β_1AR coupled to mini- G_s and $A_{2A}R$ -mini-G_s complex. a, A time-course experiment was performed to monitor the formation of active β_1AR -mini- G_s complex. The coupling efficiency (percentage) was calculated from the relative intensity of peaks assigned to β_1AR -mini- G_s coupling in the appropriate lipid-bound state. The plot indicates that mini-G_s coupling is enhanced by PtdIns(4,5)P₂ when more than two lipid molecules are bound to the receptor. Error bars represent s.d. from at least three independent experiments. b, Plot of PMF for the interaction of mini-G_s with A_{2A}R in the presence of PtdIns(4,5)P₂ (green) or PS (grey). The PMF is calculated along a reaction coordinate (Δz) corresponding to the centre-centre separation of the mini-G_s and receptor proteins along the z axis (normal to the bilayer plane). The interaction of mini-G_s with the A_{2A}R is stabilized in the presence of PtdIns(4,5)P₂ by 50 ± 10 kJ mol $^{-1}$ relative to PS. Error bars (which are <10 kJ mol $^{-1}$) are from bootstrap sampling of the PMFs and therefore represent the 'statistical' errors in estimating the well depth from a given set of simulations and PMF calculation (n = 3 independent experiments). We therefore estimate a minimum error of $\leq 10 \text{ kJ} \text{ mol}^{-1}$. c, Mass spectra were recorded for a 1:1 equimolar mix of an inactive unliganded β_1AR variant, E130W, and its unmodified active counterpart (co-purified with the agonist isoprenaline) in the presence of PI(4,5)P₂. Lipid binding occurred on both receptors, but following normalization to account for differences in ionization efficiency, a clear preference for PtdIns(4,5)P2 binding to the active receptor was observed. Bars represent mean \pm s.d.



Extended Data Fig. 7 | Detection of nanobody coupling to β_1AR . Peaks in the mass spectrum assigned to Nb6B9 binding to β_1AR to form an equimolar β_1AR -Nb6B9 complex are highlighted in orange, and demonstrate complete complex formation, implying that nanobody has a higher affinity than mini- G_s for β_1AR . n=3 independent experiments.



Extended Data Fig. 8 | Structural comparison of class A and class B GPCRs in complex with trimeric $G\alpha\beta\gamma$ complexes. The PtdIns(4,5)P₂ contacts of the $G\alpha_s$ subunit observed in molecular dynamics simulations (green spheres) are highlighted on the structures of trimeric G-protein interactions with β_2AR (PDB: 3SN6), the glucagon-like peptide-1 receptor (GLP-1) (PDB: 5VAI) and the calcitonin receptor (CTR) (PDB: 5UZ7).

Basic residues on the interface adjacent to the cytoplasmic end of TM4 are highlighted as purple spheres. Lower panels show an expanded view, highlighting the conserved pattern of PtdIns(4,5)P₂ bridging in class A GPCRs (β_2 AR and A_{2A} R (Fig. 3e)), both of which have basic residues on TM4 (Lys140 and Arg107/111) that are not present in the class B GPCRs GLP-1R and CTR.

Extended Data Table 1 | Lipidomics analysis of purified $\beta_{\text{1}}\text{AR}$

Mass (Da)	Lipid ID
824.61	PC (36:0)
716.52	PE (34:1)
742.54	PE (36:2)
714.51	PE (34:2)
772.59	PE (38:1)
752.57	PE (37:4)
793.57	PG (38:6)
807.50	PI (32:1)
833.52	PI (34:2)
865.58	PI (36:0)
835.53	PI (34:1)
863.57	PI (36:1)
861.55	PI (36:2)
891.60	PI (38:1)
758.50	PS (34:2)
760.51	PS (34:1)
786.53	PS (36:2)
1019.60	PIP (42:5)
1343.90	CL (64:4)
1371.93	CL (66:4)
1399.96	CL (68:4)



Extended Data Table 2 | Simulations run

Name	Length	Bilayer Composition
NTSR1	10 x 5 μs	POPC(95%):PIP ₂ (5%)
NTSR1	3 x 5 μs	POPC(95%):PS(5%)
NTSR1	10 x 5 μs	POPC(95%):PS(5%):PIP ₂ (5%)
NTSR1 extended	1 x 100 μs	POPC(95%):PIP ₂ (5%)
β ₁ AR (5%)	10 x 5 μs	POPC(95%):PIP ₂ (5%)
β ₁ AR (10%)	10 x 5 μs	POPC(90%):PIP ₂ (10%)
β ₁ AR (S68R, 10%)	10 x 5 μs	POPC(90%):PIP ₂ (10%)
A _{2A} R-mini-G _S	10 x 8 μs	POPC(95%):PIP ₂ (5%)

Lipids were symmetrically distributed between leaflets.

CAREERS

CONNECTIONS Dispel your doubts and convene your own colloquium **p.431**

BLOG Personal stories and careers counsel **blogs.nature.com/naturejobs**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



Retired palaeontologist Michael Wuttke takes lignite samples with a drill stick, near Darmstadt, Germany.

TRANSITIONS

Stick retirement!

Scientists who step back from full-time work can find plenty of ways to remain active in their research field.

BY AMBER DANCE

ouis Chen was technically meant to retire in 2005. The mathematician at the National University of Singapore was turning 65, the university's official retirement age. But he was only five years into his tenure as director of the university's new Institute for Mathematical Sciences, and the university

wanted him to stay on. So he remained for seven more years, stepping down in 2012. Over the next 18 months, he travelled and had knee surgery, before returning in summer 2014 to teach graduate courses for a year.

Then, in 2015, Chen's provost took him to lunch. "He told me that maybe it was time for me to leave," says Chen, who was happy to retire. But he still hasn't really left: he's at his

university office three or four times a week. "I cannot abandon my research," says Chen. "It's a passion." In July 2015, he was appointed emeritus professor, a title that comes with perks: he's eligible to apply for grants, and continues his research on probability and statistics. He maintains his e-mail address and library access and, he's delighted to say, "free parking for life".

GO YOUR OWN WAY

There are as many ways to retire as there are scientists; there's no right or wrong path. Many researchers wish to continue their academic lives in one way or another. The emeritus title can allow scientists to keep laboratory or office space or apply for grants; associated privileges vary widely. However, research funds probably won't flow as generously as they used to, and emeriti typically downsize their research space and teams. Some retired scientists turn to other projects, such as writing books or doing charity work. The key to a fulfilling retirement, say those who are pleased to have stepped down from full-time work, is to line up positions and projects, and to prepare for the emotional toll that the transition can take.

Worldwide, the ranks of those aged 60 or older are expected to rise. The United Nations, for example, predicts that by 2050, 21% of the world's population will be at least 60 years old, up from 10% in 2000. Among scientists, in particular, the average age is climbing. In the United States, scientists' average age rose from 45 to 48.6 between 1993 and 2010, and it is expected to climb further. The trend is similar in Europe.

National rules on retirement vary widely. In Sweden, for example, it is mandatory at age 67; in South Africa, at 65. The United States and Canada have no mandatory requirement age. Although data on active retirees and emeriti are scarce, a 2014 survey of retired medical professors from 20 countries found that many continued to teach, and that more than 40% had published at least one paper or book in the previous year (N. G. De Santo *et al. QJM* 107, 405–407; 2014).

Researchers who are nearing retirement should start preparing for it as early as possible, advises Amy Strage, assistant vice-president for faculty development at San Jose State University in California. There could be many options to research and decisions to weigh. For example, at some universities where retirement is a choice, faculty members can take advantage of phased retirement plans. That means they can wind down their research while working part-time, so long as they commit to a date

▶ for full retirement within a few years. Those required to step down from their positions at a certain age might be able to arrange unpaid positions, or jobs in countries with a higher retirement age.

Some retired faculty members gain emeritus status, although the meaning of that title varies widely between institutions and nations. At some universities, it's granted pro forma to retiring full professors. At others, it's an honour bestowed only on pre-eminent researchers. "It is retirement with distinction," says Kimberly Read, assistant director for the Florida Center for Inclusive Communities at the University of South Florida (USF) in Tampa. Read researched retirement and emeritus issues, focusing on the oral history of an emeritus professor, for her 2016 PhD thesis at USF.

Emeritus is the final rung on the academic trajectory from assistant professor to associate to full. Obtaining this ultimate promotion is often much like gaining those earlier ones, with a committee evaluating a person's research or service contributions to the university, and administrators approving a decision to award the honour.

For some, the emeritus title is a final feather in their academic cap as they head through the door. Others take it as a commitment to further engagement with the university. "You want to continue to help the department," explains Dean Martin, an emeritus professor of chemistry at USF and Read's research subject. Every morning, he comes to his office, where he does research and publishes papers,

LASTING BONDS

Keeping in touch

Full-time researchers interact daily with colleagues and students, but retirees risk losing that sense of community. Organizations can help to restore it. Here are a few examples.

- The European Association of Professors Emeriti welcomes retired professors from all European universities, as well as corresponding members from abroad. From 68 founding members in 2016, the organization has grown to nearly 200.
- The US-based Association of Retirement Organizations in Higher Education lists about 100 such communities, often associated with a particular university or other academic organization.
- The Emeritus College of Arizona State University (ASU), in Tempe, welcomes emeriti of ASU and associate members from elsewhere. Member activities include memoir classes and helping students to prepare for an international science competition. A.D.



Pharmacologist Edith Sim enjoys her retirement.

mentors students, edits the departmental newsletter and raises funds for the department.

Achieving emeritus status gives retiring professors a lasting connection with their university. They might or might not be given their own office, but they will typically have access to resources such as the gym, library and e-mail. They might also be given admittance to emeritus associations, which provide camaraderie (see 'Lasting bonds'). They will not, however, receive a salary.

Access to grant money varies from one country to another, but maintaining emeritus status and a university affiliation is often enough to make a retired researcher eligible, at least to apply. "Since I retired I'm busier than ever, writing papers, travelling to meetings and giving talks," says George Ellis, 79, an emeritus mathematician at the University of Cape Town in South Africa. "The main issue is funding."

For several years, he held on to a small grant of 100,000 rand (about US\$7,500) from the South African National Research Foundation. The amount allowed him to attend overseas conferences and invite researchers from other nations to visit and collaborate. But cuts at the foundation have caused the grants to dry up. Now, he plans to attend conferences only if the hosts pay for his trip.

Fortunately for Ellis, his research requires little funding and few resources. "They can't keep one from thinking and reading and writing," he points out. With his emeritus status, he's able to keep an office in his department, and he continues to work with colleagues and students.

Some researchers manage to keep a lab going. Martin has continued to win grants well into his retirement. But he is an outlier in terms of the amount of work he does in his retirement and, at times, this has generated confusion. His

grant funds are deposited in a research bank ≥ account — but more than once, the university assumed he was inactive and transferred those funds into central accounts. (His dean and department chair restored the money.)

Many retired scientists, however, can't maintain the lab space and funding that they did as active researchers. They might not want to continue competing for grants, and the space might be needed for new faculty members. That doesn't mean that retired researchers can't make academic contributions, says Strage. They might shift from bench science to less-space-intensive activities such as giving speeches, guest teaching or reviewing manuscripts, she explains.

Others find new lab space. Michael Wuttke, a vertebrate palaeontologist, engineered his own post-retirement research position. In 2015, aged 65, he left his job at the General Directorate for Cultural Heritage Rhineland-Palatinate in Mainz, Germany. But he had started seriously considering his next steps a few years earlier. He set up a position as 'designated volunteer' at the Senckenberg Research Institute and Natural History Museum in Frankfurt, and since 2015 he has been working with specimens from fossil sites such as the Messel Pit, a disused quarry near Frankfurt, where he did his PhD research. Wuttke has access to the same resources and scientific equipment that employees have. He expects to publish soon on a previously unknown species of frog, whose fossilized remains were discovered at the site.

TEAM SPIRIT

Scientists can also remain active in research by continuing to correspond with the team they once led. Stem-cell physician Outi Hovatta retired from the Karolinska Institute in Stockholm, at the age of 70. Although Sweden's retirement age is 67, she was able to stay on for three more years as a senior faculty member — provided she funded her own salary from grants. At 70, she was happy to return to her family home in Helsinki.

She passed the research group on to a colleague, but stays active as a professor emerita. Now 72, Hovatta continues to correspond with her Karolinska colleagues. Including her name on grants helps them to obtain funding, and she comments on draft publications.

Edith Sim, 67, a retired pharmacologist and emerita professor of the University of Oxford, UK, and Kingston University London, has also kept up her research without a lab. She's writing papers based on unpublished data that she had collected earlier, and on fresh data from collaborators. She recently published a book, co-authored with a former student, on how certain enzymes affect a person's response to

Sim is also involved in charitable work. This includes running a Saturday morning programme to give teenagers a taste of what science is like before they commit to studying it further. For example, the teens in her pilot

programme at Kingston played with wind tunnels and microscopes and interviewed astronauts. Sim is also a trustee with the Daphne Jackson Trust, a UK charity that helps scientists who have had a career break to return to research.

Of course, scientists don't always find retirement easy. Nancy Schlossberg thought it would be "a piece of cake". In 1997, at the age of 68, and after 24 years as a counselling psychologist at the University of Maryland in College Park, she became an emerita and headed for Sarasota, Florida. There, she hoped she'd find ways to write or get speaking invitations. But it wasn't that simple. "I get to Sarasota, and there I am without a purpose," she recalls. "I was shocked."

Schlossberg had to create her own opportunities. Because her professional expertise was in life transitions, she decided to study retirement. That work led to the first of three books that she has penned on the topic since 'retiring': *Retire Smart*, *Retire Happy* (American Psychological Association, 2003).

Through her post-retirement research, Schlossberg has worked out why it didn't feel good to be suddenly purposeless in sunny Florida. Those who have retired, she learned, must shape a new identity. "That transition process, even if it's something you

"Since I retired I'm busier than ever, writing papers, travelling to meetings and giving talks." wanted, can be very unsettling," says Schlossberg.
"But the most important thing tied to your identity is your sense of purpose. That's what gives you the

reason to get up in the morning." Of course, that sense of purpose or identity needn't be related to previous professional activities — but Schlossberg's interviews with retired researchers indicate that they need something to define their lives.

"You do feel, a little bit, that you might be kind of sidelined," says Chen, who has noticed that he receives fewer invitations to present at or to organize talks and conferences. But he's not terribly bothered. "I think you have to accept this," he says.

And in any case, Chen notes, he now has time to lunch with old friends whenever they call, and to rekindle old hobbies: singing, and playing the recorder and cello.

Sim also struggled a little at first. As a full-time scientist, she'd found solace in her garden. Once she retired, gardening no longer offered the same sense of escape. It took time to rediscover the joy of tending the plants. Today, she says, retirement feels good. "Now that I've got used to it," says Sim, "it's a very nice way to live."

Amber Dance is a freelance writer in Los Angeles, California.

COLUMN More than a meeting

Convene a colloquium, says Francesco Sciortino.

rganizing a scientific conference can be a daunting prospect. You know that it could offer exceptional career benefits by boosting your network and helping you to develop those famous soft skills: communication, teamwork and time management. But you might think the process involves unacceptable levels of stress, complications to your unpredictable schedule and even more delays to that unfinished project.

Still, you should consider the option. You'll refine skills that are not necessarily innate and that you'll need in any job. Why not hone them in the setting of an enthusiastic student group?

I became involved in student activism during my high-school days in Italy, before I moved to the United Kingdom in 2010 to study physics. As an undergraduate at Imperial College London, I joined student associations to meet like-minded people and to get a taste of a variety of research fields. I set up tours to my department's laboratories and found the gratitude of other students to be extremely rewarding. Through the Imperial College Physics Society, I also co-organized a number of trips, some of which later inspired me to pursue a PhD in plasma physics — none more so than our visits to the Culham Centre for Fusion Energy near Oxford in 2013, 2014 and 2015.

A different chapter began in August 2014, when I and six others joined together to found the Italian Association of Physics Students (AISF). Since then, our group has grown to more than 1,000 members in Italy and has become one of the most active in the International Association of Physics Students (IAPS). We have organized public lectures, lab tours and outreach events, offering simple demonstrations to school groups of all ages and engaging in the International Year of Light celebrations in 2015, which aimed to highlight the importance of light and optical technologies. Since then, the AISF has also set up annual visits to Italy's Gran Sasso National Laboratory in Abruzzo, the European Gravitational Observatory near Pisa and other leading research facilities. The Italian Conference of Physics Students has become our key annual gathering, bringing together more than 100 students from institutions nationwide in a different city each year.

In 2015, one year after we founded the AISF, we submitted a bid to host the 32nd International Conference of Physics Students (ICPS). It sounded a little over-ambitious at first, but we demonstrated that our association could raise the necessary funding and institutional



support. It could hardly have gone better. In August 2017, the ICPS took place in the Italian city of Turin with 450 participants from 44 countries, and included almost 200 talks and posters from university students of all levels. I was part of an outstanding team that helped to exhibit the best of Italian academic research, the wonders of our national cuisine and local artistic treasures. Our programme included trips to the Turin Astrophysical Observatory, Sacra di San Michele Abbey and traditional wine cellars.

Organizing student events shapes how you collaborate with people. I discovered what kind of team player I am. I learnt that balanced group dynamics, rather than individual herculean efforts, best foster motivation, enthusiasm and effectiveness. I've always wanted my impact to exceed my direct reach, and so connecting with others who could carry my efforts forward was essential. Seeing other people independently repeat events that I initiated has been extremely rewarding.

I started out with pragmatism, but little understanding of the art of compromise. That's now been forced into me by countless online meetings, most recently as part of a committee to reform the regulations of the IAPS. The international setting of these efforts also gave me chances to travel, practise languages and gain exposure to fund-raising. I've developed important friendships and boosted the competitiveness of my PhD applications, which in turn brought me to the United States.

Joining student associations, organizing events all over Europe and becoming part of a community of enthusiastic young scientists has helped me to go beyond lecture halls, research labs and supervisor meetings. The skills that I gained have given me the freedom to enjoy much more of my own scientific career. ■

Francesco Sciortino is a PhD candidate in plasma physics at the Massachusetts Institute of Technology in Cambridge.

PAPA BEAR

Cold comfort.

BY KURT PANKAU

mages flash in my head. Evacuations. Fires. Bodies. The long winter. Cold. I'm on a mattress on the floor, one arm draped over a woman I don't recognize. It's dark, but I can see wheel wells. So it's not the floor — the back of a truck,

The sleeping woman in front of me is facing away, but her face looks young, maybe early forties. She's wrapped in several layers of ragged clothing. And so am I, I realize. I reach a hand up to feel my face and find a thick beard. I never wore a beard. I don't have the slightest idea how to maintain one.

I sit up and place a hand on the window. There's a curtain, but through the thin cloth I can feel the freezing pane of glass. I pull aside the curtain and I can see a foot of snow reflected in the moonlight. The sky is dark purple. I grope around for the door.

Where am I?

"Albuquerque, I think," says the woman. "Come back to bed, Papa Bear." I don't recognize her voice.

"I'm just going to take a leak," I say. I need to get out of here.

"We emptied your bag three hours ago," she said.

I reach a hand instinctively to my middle and feel a twitch in my abdomen as my fingers nudge a plastic line attached to the stoma near my ... How do I know what a stoma is?

"Come back to bed."

"Who are you?" I ask.

"It's me," she says. "Lisa."

"Lisa's long dead," I say. How do I know that? Dear God, Lisa. When did she die?

"Dammit," says the woman.

"Why did you say that?" I ask, my anger rising. "Why would you pretend to be my wife?'

"I wasn't pretending to be your wife." The woman makes a snorting noise. "I never know what you're going to remember."

"Remember?"

"During your spells."

"Spells?" I fumble for the door.

◇ NATURE.COM Follow Futures:

@NatureFutures **1** go.nature.com/mtoodm

"Please don't," she says. "If you let in the cold air, we may

"Then turn on the

engine," I say, moving to the front of the van. There's no key, but the panel under the steering wheel has been ripped off and wires dangle, exposed. The passenger's seat is filled with gasoline cans and jugs of water.

Scraps of paper with a tangle of notes in my handwriting. My fingers are trembling. Letters to myself. Pieces of a puzzle I can't quite assemble.



It's like the past ...

"Like the past is running away from you?" I feel hands on my shoulder and her nose on the back of my neck. The touch is famil-

"How did you know —?" I start to ask.

"You were thinking out loud," says the woman. "You do that."

"Oh," I say.

"I'm cold, Papa Bear," she says. "Please come back to bed. I'll explain in the morning."

Explain it to me now," I say, turning to face her. "Tell me everything."

'You'll never get back to sleep," says the woman. "There's too much. Come back to bed. I want you to sing to me."

"Who are you?" I ask.

"I'm your wife," she says.

"Lisa's dead," I say.

"So is my first husband. So are a lot of people." She wraps her arms around my neck and presses her lips to mine. The hairs of my beard rub against my chin and tickle my face. The kiss is brief, but I know those

lips. She pulls back and rests her forehead against mine.

"Tell me your name, at least," I say.

"Jennifer," she says. Her voice is quavering. Oh no, I've done something wrong.

"No, you haven't," she says, shaking her head gently.

"I'm just ... I'm having trouble," I say. My voice is hoarse and raspy.

> "I know, Papa Bear," she says, sniffling. "I promise, I'll tell you everything in the morning. Just come back

"Why are you crying?" I ask.

"Because a long time ago, the world ended," she says. "The world ended and you saved us and I c." love with you. And because nobody else gives a damn about you, even though you saved all our lives" — she swallows hard — "because you get confused sometimes."

I look out the window. There are a dozen other cars around us. Like a caravan. Beyond them, lean-tos and the skeletons of small structures. They're building something. No, they're rebuilding. Why? We need to stay mobile. Don't we?

"I know where this conversation is heading and I can't do it again, not tonight," she says. "I just can't. Please come to bed. I promise, everything will be better in the morning.

Just come back to bed." Tears glisten in the moonlight. In this cold, they must sting her cheeks.

I nod. She pulls me back to the mattress. She lies down and holds my arm tightly to her chest. Her face flashes through my mind. I can feel the weight of a thousand memories hidden behind a fog.

"I love you, Papa Bear," she says. Her voice is almost a whimper.

"I love you too, Baby Bear," I say, more from instinct than memory.

Her gentle sob is interrupted by a burst of tearful laughter. I guess I've remembered something important. I pull her closer to me, to shut out the cold, the emptiness, the void in my head where the past should be. I close my eyes.

Everything will be better in the morning.

Kurt Pankau lives with his family in St Louis, Missouri. He loves board games, dad jokes and stories about time travel. He tweets at @kurtpankau and occasionally blogs at kurtpankau.com.